The importance of socio-economic status in determining educational achievement in South Africa

STEPHEN TAYLOR AND DEREK YU

Stellenbosch Economic Working Papers: 01/09

KEYWORDS: SOUTH AFRICA, SOCIO-ECONOMIC STATUS, EDUCATION, EDUCATIONAL ACHIEVEMENT, EDUCATIONAL INEQUALITY, ECONOMIC DEVELOPMENT
JEL: I20, I21, I30, O15

STEPHEN TAYLOR
DEPARTMENT OF ECONOMICS
UNIVERSITY OF STELLENBOSCH
PRIVATE BAG X1, 7602
MATIELAND, SOUTH AFRICA
E-MAIL: STEPHEN@SUN.AC.ZA

DEREK YU
DEPARTMENT OF ECONOMICS
UNIVERSITY OF STELLENBOSCH
PRIVATE BAG X1, 7602
MATIELAND, SOUTH AFRICA
E-MAIL: DEREKY@SUN.AC.ZA

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

The importance of socio-economic status in determining educational achievement in South Africa

STEPHEN TAYLOR AND DEREK YU

ABSTRACT

The needs to find ways of lifting people out of poverty and to transform the existing patterns of inequality in South Africa are high on the country's development agenda. Much hope is often vested in education as an opportunity for children from poor households to overcome the disadvantage of their background and escape poverty. The logic of this is often conceived of in terms of the human capital model, according to which education improves an individual's productivity, which in turn is rewarded on the labour market by higher earnings. However, there is a circularity in the relationship between socio-economic status (SES) and education, in that it is well known that a student's SES has an important influence their educational achievement.

Drawing on data from the recent Progress in International Reading Literacy Study (PIRLS 2006), this paper investigates the extent to which SES affects educational achievement in the case of South Africa, and moves on to consider the implications of this for the ability of the education system to be an institution that transforms existing patterns of inequality rather than reproducing such patterns.

Keywords: South Africa, socio-economic status, education, educational achievement, educational inequality, economic development
JEL codes: I20, I21, I30, O15

Introduction

The recent release of the PIRLS[1] 2006 results has added to the growing body of evidence suggesting that South Africa's school system is seriously underperforming. South Africa's mean reading score is the lowest out of the 40 participating countries in PIRLS 2006. This result is in line with a similar international survey – TIMMS[2] 2003 – where South Africa recorded the lowest mean scores in both mathematics and science, out of the 50 participants. These results cast doubt on the ability of the South African school system to play an effective part in addressing the country's developmental needs, not least of which is a transformation of the vastly unequal distribution of wealth and income.

The relationship between education and this developmental goal of improving the distribution of income contains an element of circularity. Although education is often looked to as an opportunity for children to overcome the disadvantage of social background by placing themselves on an equal footing with others upon entering the labour market, it is well known that the Socio-Economic Status (SES) of children's families has a significant influence on their educational achievement. And of course educational achievement is a good predictor of performance on the labour market, thereby completing the circle. Instead of transforming patterns of inequality within society, an education system may actually reproduce such patterns.

This paper explores the influence of SES on educational outcomes in South Africa and considers some implications for social mobility. In Section 1, the relevant concepts of economic development and SES are defined, and the way in which they interact with education is conceptualised. Section 2 provides an introduction to the PIRLS 2006 data as well as a preliminary overview of South Africa's performance. In Section 3 the technique of constructing SES gradients is applied in order to investigate the relationship between SES and educational achievement in South Africa. This analysis is extended in Section 4 with a more comprehensive multivariate analysis. In Section 5 the implications of the strong relationship between SES and educational achievement for the prospects for social mobility are considered. The paper concludes with a discussion following from the major findings of the paper. This discussion includes some implications for policy.

---

[1] Progress in International Reading Literacy Study
[2] Trends in International Maths and Science Survey

1. Education, Economic Development and SES

Before considering the links between education and economic development the latter concept needs to be defined. . Meier (1995: 7) describes economic development as a process of long-term *per capita* growth that leads to qualitative improvements throughout the social system. This definition captures the notion that growth in *per capita* output sustained over a long period of time is often the major driving force behind development. Moreover, Meier's definition emphasises that development is more than just growth. It includes qualitative improvements throughout society.

Focussing for the moment on the "growth" component of development, there are strong theoretical reasons to expect education to contribute to economic growth. Education raises the human capital of the labour force, improves the innovative capacity of the economy and facilitates the transmission of new knowledge and technologies. Indeed there is a substantial literature that explores the inclusion of education in growth regressions, but this is not the present focus.[3]

In trying to unpack the more qualitative component of development, it can be said that to some extent the developmental agenda of an economy is set by its particular needs. Therefore in highly unequal societies such as South Africa, addressing the distribution of income and wealth is a core aspect of the development challenge. Much hope is often placed on education as an institution of transformation. The mechanics of this is usually conceived of in terms of the human capital model, according to which education improves an individual's productivity, which in turn is rewarded on the labour market by higher earnings. In this way it is hoped that education can give an opportunity for children coming from poor socio-economic backgrounds to perform well on the labour market, thus overcoming their disadvantaged background.

An increasingly influential and far-reaching conception of development is offered by Amartya Sen. In his approach, development is a "process of expanding substantive freedoms that people have." (Sen, 1999: 297) Freedom consists in having the "capabilities" to live the sort of life one has reason to value. These capabilities range from basic survival abilities to the ability to

[3] For example, the seminal contribution of Mankiw, Romer and Weil (1992), where human capital is introduced into the Solow-Swan model.

function well in society. Sen emphasises that there is a complex interconnectedness amongst these freedoms and capabilities. For example, illiteracy and under-nourishment are often results of low income. And yet conversely, education and good health are important determinants of income (Sen, 1999: 19). This interconnectedness amongst freedoms may explain how improving access to education can have a limited impact on well-being if other "unfreedoms" persist. This motivates the major research question of this paper: To what extent does SES determine children's educational performance and thereby constrain economic development in South Africa?

Any account of the influence of SES on education should take stock of the seminal work of the Coleman Report of 1966. James Coleman was commissioned to investigate the inequalities of educational opportunity in the United States, with the assumption that race would be the major focus. However, Coleman's findings were not entirely as expected. The disparities in spending on black and white education were far less substantial than expected. Neither did funding turn out to be a very good predictor of educational achievement. Instead family background and SES was found to explain much of the patterns in achievement. Moreover, Coleman found that a more important resource than school funding was the effect of school peers, in particular the socio-economic backgrounds of peers (Kahlenberg, 2001).

Research subsequent to the Coleman Report has widened the consensus regarding the importance of SES in determining educational outcomes. It is well known that family SES is a major determinant of educational attainment (e.g. Filmer and Pritchett, 1999) and of the quality of schooling likely to be received (e.g. Barro & Lee, 1997). The intention in this paper is to investigate just how strongly SES determines educational achievement in South Africa – a highly unequal society. This investigation feeds directly into the broader question as to whether education systems transform or reproduce patterns of inequality. Or, more specifically, to what extent is the schooling system in South Africa transforming or reproducing existing patterns of inequality?

Before plunging into a quantitative analysis of the influence of SES on educational achievement in South Africa, it is necessary to establish a definition of SES and then to consider some of the ways in which one might expect SES to influence educational achievement. Willms (2004: 7), quoting Mueller and Parcel (1981), defines SES as the "relative position of a family or individual on an hierarchical social structure, based on their access to, or control over, wealth, prestige, and power." In economics, where the intention is often measurement, SES tends to be conceived of in terms of its proxies, such as income, education or occupation. In sociology, which is where the concept emanates from, SES is very much conceived of in terms of societal rank, prestige and position (Bullock and Stallybrass, 1982: 599).

A broad conception of SES fits well with Sen's "capabilities" approach to development. Reading ability, which is the measurable educational outcome used in this paper, perhaps more so than Maths or science, contributes strongly to the ability of people to function well in society. In terms of Willms's definition of SES, including the ability to gain access to power and prestige in society, literacy is important beyond merely its role on the labour market. According to Willms (1997: 22), it is important for being included in a culture, and for expanding social relations and networking which facilitate access to positions of influence and power in society.

But of course access to money income is also crucial in a capitalist economy. This is where the literature dealing with labour market returns to education comes in. Following the work of Jacob Mincer, an extensive literature using earnings functions has emerged based on the logic of the human capital model. These functions regress earnings on a variety of personal characteristics such as age or experience, gender and, importantly, education. Studies have indeed consistently shown that more years of schooling are associated with higher earnings over an individual's lifetime. Hanushek and Woessmann (2007) extend this literature by adjusting for the quality of education in specifications of earnings functions. They find that educational quality has a strong influence on individual earnings over and above the quantity of schooling.

In South Africa, the hierarchical structure of society, including access to wealth, prestige and power, was constructed to be on the basis of race through decades and even centuries of institutionalised inequality. This was achieved by placing restrictions on where people could live, the type of education they had access to and the work occupations they had access to. Thus history has ensured that SES is distributed along racial lines. It is therefore tricky and even ill advised to attempt to untangle race and class in the case of South Africa. For example, one area where family background characteristics are distributed by both race and SES is family structure. Non-traditional family structures, such as the phenomenon of skip-generation households, are known to be most prevalent in low SES households. In South Africa, this has a strong racial dimension. In 2006, 31% of black children between the ages of ten and twelve lived in a household with neither parent present, 41% of black children lived with a single parent and only 28% lived with both parents present. In contrast, 80% of white children and 89% of Indian children between ten and twelve lived with both parents present (calculated from the General Household Survey, 2006). Research has shown that this aspect of family background has a significant effect on educational outcomes. For example, Anderson (2000: 12-13) finds that family structure strongly affects the current enrolment status of students, their highest grade completed as well as the number of years delayed in school if still enrolled.

There are a number of channels through which SES can be expected to affect educational outcomes, and for which there is evidence in the literature. The most direct way this happens is through home support. Bearing in mind that parental education is one component of SES, it is likely that better educated parents can directly benefit their child's education by starting the education process during pre-school years. Lee and Burkham (2002), in their book called "Inequality at the Starting Gate", find significant disparities in the cognitive ability of children upon starting school associated with SES background. Once schooling is underway better educated parents are able to offer direct support such as by helping with homework. Moreover, they have easier access to information that will help with their children's health, social and emotional well-being, all of which feed into educational achievement. Anderson, Case and Lam (2001: 6) consider that high SES and better educated parents also may indirectly advantage their children's educational achievement by being able to live in neighbourhoods where there are better schools, or by being able to choose to send their children away to good schools, perhaps at a financial cost. Another mechanism to consider is that better educated and high SES (including social prestige) parents are more likely to get involved in the school community, thus increasing the sense of accountability school staff feels towards the parents. This mechanism is likely to be strongly at work in schools where the parent body contains predominantly high SES parents. In such schools the accountability structures will be well developed and contribute to school quality.

Neighbourhood effects provide another mechanism through which the concentration of similar levels of SES affects education. This is particularly acute in areas of concentrated poverty. Contagion theories of poverty emphasise how the sorts of problems that are associated with poverty spread through society much like a contagious disease when there is a high concentration of poverty in a community. Thus the problems poor households are generally vulnerable to are amplified through the concentration of poverty in the neighbourhood. According to the New South Wales Department of Education and Training (2005: 14), some of these neighbourhood problems include unsafe streets, a lack of economic opportunities, the absence of positive role models and a high concentration of non-traditional family structures. To this one can add that poor neighbourhoods tend to foster a general attitude of hopelessness and low self-efficacy.

A related but distinct mechanism is within-school peer effects. As Coleman’s research demonstrated, the social composition of schools was a more important determinant of educational achievement than school spending (Kahlenberg, 2001). According to the Coleman Report, greater integration of students from different socio-economic backgrounds was more conducive to achievement. When a high concentration of low SES students exists, attitudes that are anti-school and disruptive tend to be prevalent, and discipline becomes exasperatingly hard to maintain.

A final mechanism to consider here is that schools with mostly low SES students generally suffer from resource shortages. In the case of South Africa, government spending on education was vastly unequal across the race groups under the previous regime. Since the political transition there has been significant progress made towards greater equity on educational spending. However, the backlog is extensive and many formerly disadvantaged schools remain subject to infrastructural and resource shortages. An important shortage in this regard that is somewhat less tangible is that the better teachers tend to be concentrated in the wealthier schools. It is important to note also that there is an increasing realisation amongst economists of education as well as policy-makers that increased spending on poor schools is not translating into improved educational outcomes.

This theoretical and literature review provides a backdrop for the empirical analysis of the effects of SES on reading scores in South Africa, using the PIRLS 2006 dataset, in the sections to follow.

2. Data and Methodology

2.1 PIRLS 2006

In 2006 the International Association for the Evaluation of Educational Achievement (IEA) conducted the second round of the Progress in International Reading Literacy Study (PIRLS). The first round was conducted in 2001. The chief objective of PIRLS is to provide information about reading achievement in primary schools that will be relevant for policy and instruction. Testing was done in 40 countries, including Belguim with two education systems, and Canada, with five provinces that were analysed separately. Therefore, there were 45 participants in total. Testing was done on students in the fourth grade, with the exceptions of Luxembourg, New Zealand and South Africa, where testing was done on fifth grade students, and Slovenia, where testing was done on students in both grades 3 and 4. The reason given by IEA for testing fifth grade students in South Africa was the challenging context of having multiple languages of instruction.[4]

In addition to reading scores, PIRLS collected a wide range of information on student home background and on various school processes that feed into the process of learning to read. There were six questionnaires in total – a student questionnaire, a home questionnaire, a teacher questionnaire, a school questionnaire, a curriculum questionnaire and the reading test booklet. This information allows us to investigate the impact of SES on reading performance, one aspect of educational achievement.

The reading scores were calculated using average scale scores. This involved setting a scale average score of 500 across the countries and a standard deviation of 100. [5] For the purposes of

[4] It would perhaps have been better to test grade 4 students in South Africa and treat the issue of multiple languages of instruction as one factor feeding into educational performance, rather than to somehow attempt to build this into the design of the survey.

[5] A fairly complex procedure was followed in the calculation of the overall reading scores. PIRLS wanted to test students on 126 assessment items. However, the estimated time this would take a student to complete is 400 minutes. In order to deal with this, items were divided into 10 test blocks. These 10 blocks were then distributed across 13 test booklets – 2 blocks in each booklet – with as many different combinations of blocks as possible. Using Item Response Theory, the scores were then imputed as if the students had answered all 126 items. Of course this leads to some degree of error. Therefore, in order to provide researchers with some indication of the bias this imputation causes, 5 different plausible values were imputed. The reading scores as presented and used for analysis in this paper are calculated by taking the average of the 5 plausible values.

comparison and analysis, four points on the reading score scale were selected as international benchmarks:

Low International Benchmark:	400
Intermediate International Benchmark:	475
High International Benchmark:	550
Advanced International Benchmark:	625

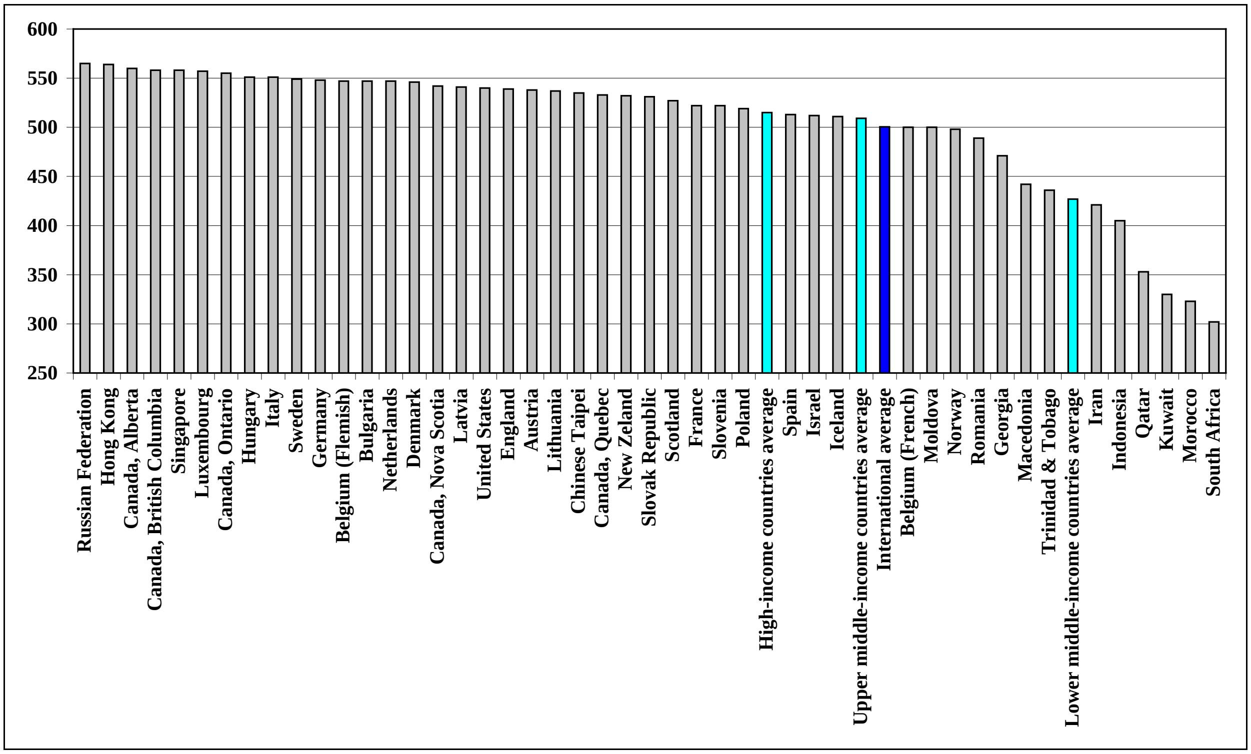
Although this paper focuses chiefly on PIRLS 2006, some comparison is made using the three waves of TIMSS (1995, 1999 & 2003) and SACMEQ[6] II (2000). These form part of the same class of international surveys of educational achievement as PIRLS. The TIMSS surveys were also conducted by the IEA and examined mathematics and science achievement. The SACMEQ surveys (1993 and 2000) tested students in mathematics and reading.

2.2 Introduction to the results of PIRLS 2006

Interestingly, the top-performing participant in PIRLS 2006 was Russia with a national average reading score of 565. The most concerning result from the perspective of this paper, is that the worst performing participant was South Africa, with an average score of 302. The mean scores for all the participants, including the mean scores for the high income, upper-middle income and lower-middle income groups of countries, are presented in Figure 1. [7] Note that South Africa is classified as an upper-middle income country.

[6] Southern African Consortium for Monitoring Educational Quality
[7] The table of overall results is presented in greater detail in Appendix A.

Figure 1: Mean overall reading achievement score



An alternative way of describing the overall results is presented in Figure 2. This shows the percentage of pupils in each international benchmark category. Singapore and Russia have the greatest proportion of students in the top international benchmark category (19.4% and 18.9% respectively). In contrast, the countries with the largest percentage of students failing to reach the Low International Benchmark category are South Africa (77.8%), Morocco (74.3%) and Kuwait (71.8%).

Figure 2: Percentage of pupils in each international benchmark category

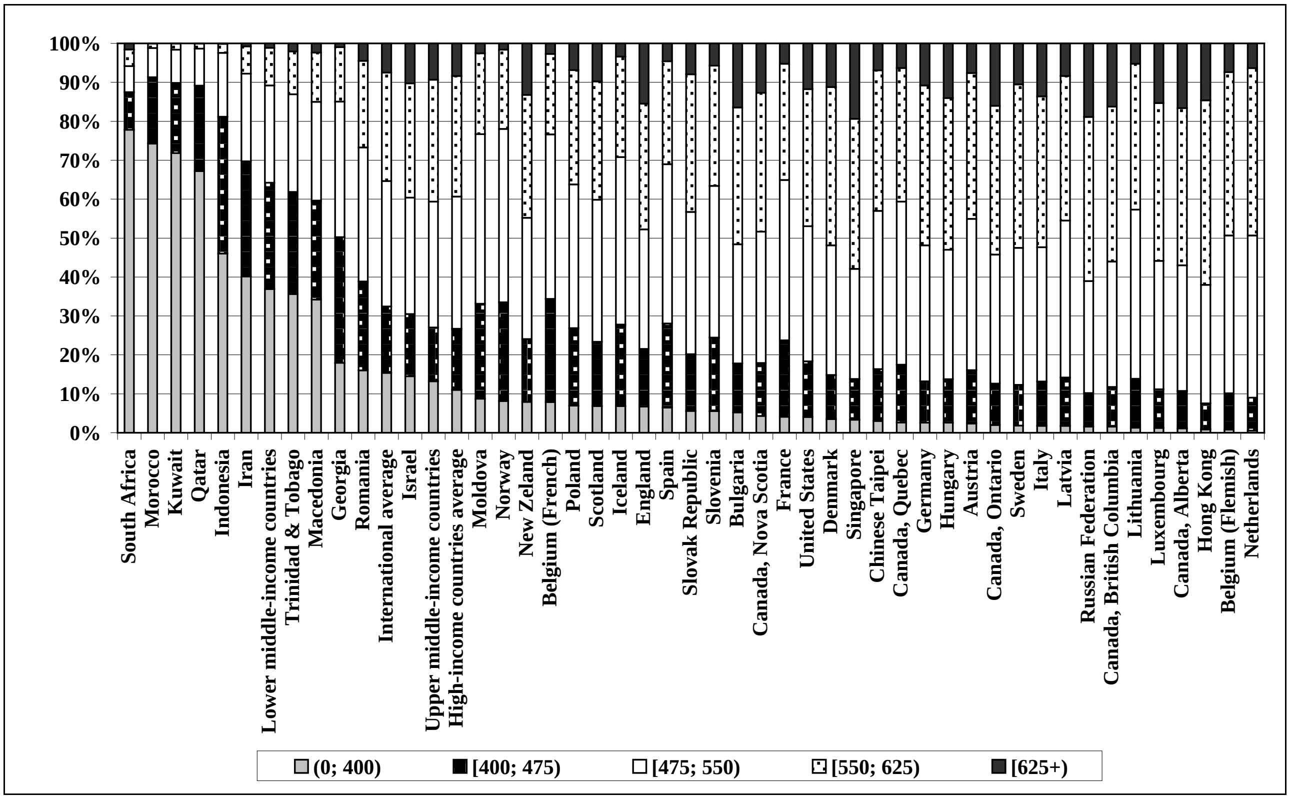
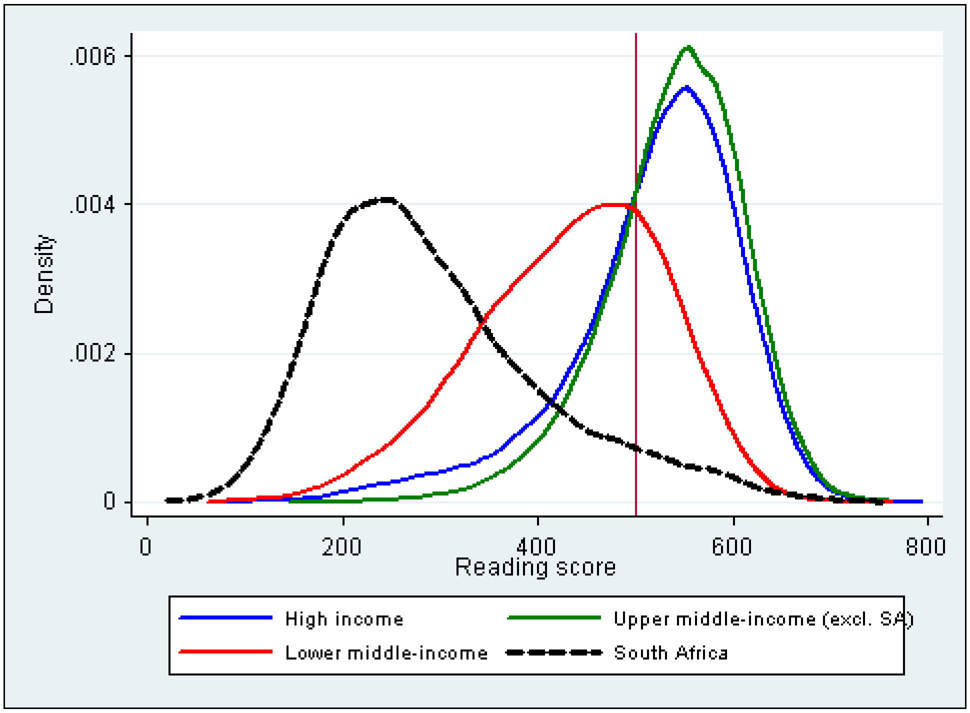


Figure 3 presents Kernel Density Curves of the reading scores of South Africa and the three different income groups of countries. [8] As the figure shows, the distribution of reading scores for South Africa lies far to the left of the other distributions, even in comparison with the lower-middle income group.[9] The other noteworthy difference with South Africa's Kernel Density Curve is that it is much flatter, especially on the right hand side of the distribution. This is indicative of the great variance of reading scores and thus of the high level of educational inequality. Indeed South Africa has the highest variance and standard deviation (136) of all the participants in PIRLS 2006.[10]

[8] See Appendix B for a brief explanation of kernel density curves.
[9] It may seem surprising that the Kernel Density Curve for the high income group lies slightly to the left of that for the upper-middle income group. This result is largely attributable to the poor performances of Qatar and Kuwait – both high income countries.
[10] Recall that the overall average standard deviation was set at 100.

Figure 3: Kernel density curves by country income groups

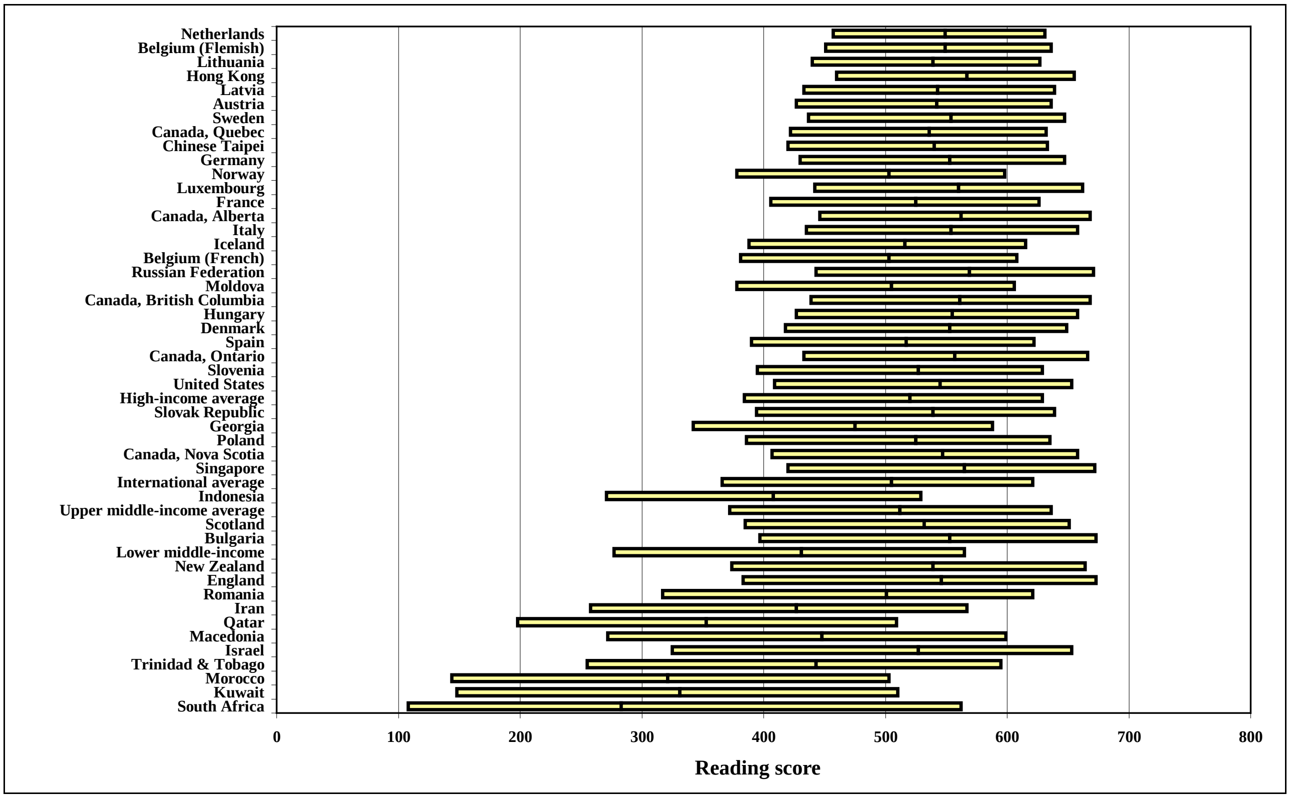


Note: The thin red line marks the international mean score of 500.

The great variance of reading scores for South Africa is alternatively demonstrated in Figure 4, in a visually accessible manner. The countries are arranged by the size of the difference in reading scores between the 5th and 95th percentiles. As expected, Figure 4 shows the South African distribution far to the left relative to the other countries, but also with the greatest difference between the 5th and 95th percentiles. Disturbingly, the reading score at the 5th percentile for South African students is 108. Bearing in mind that the international average is set at 500, and that the Low International Benchmark value is 400, and that many of the test questions were in multiple choice format[11], this result would indicate that there is group of students enrolled in the South African school system that is effectively illiterate.

[11] Test questions were either in multiple choice format or constructed response format.

Figure 4: 5th percentile, 50th percentile (median) and 95th percentile of performance



Note: the countries are arranged in ascending order by the (95th percentile – 5th percentile) difference

The disturbingly poor performance of the South African students in PIRLS 2006 together with the great variance in the results, contributes to the relevance and urgency of an investigation into the role of SES in determining these outcomes.

2.3 Measuring SES in surveys such as PIRLS, TIMSS and SACMEQ

This paper uses PIRLS 2006 as its major dataset. In Section 3 there is some comparison with the results obtained when applying similar techniques to data in TIMSS and SACMEQ. When examining SES in such datasets an approach to measuring SES needs to be decided upon.

One might initially suppose that the absence of any information on household income or expenditure in these surveys is a major obstacle to measuring SES. However, there are several problems associated with measuring income in household surveys. Firstly, there tends to be a high incidence of non-response which is usually not missing at random, creating a bias in the results. Secondly, measurement error often occurs as a result of households having multiple sources of income. In particular, poor households receive a significant amount of income in kind. Attempts to collect income data in a way that is sensitive to these issues are time-consuming and costly. Moreover, income is rather subject to short-term fluctuations and is therefore not always a good measure of long-term wealth or SES. For the purposes of this paper we are interested in a long-term measure of household SES, as this is most likely to be what drives educational achievement.

For these reasons, consumption or expenditure is often regarded as a better proxy for long-term SES than income. However, the collection of expenditure or consumption data also carries a fairly heavy burden of time and therefore cost. Furthermore, Filmer and Pritchett (2001: 12) point out that in the real world consumption smoothing happens on the basis of imperfect foresight and imperfect capital markets. This detracts somewhat from the case for expenditure data being a good proxy for long-term SES.

An increasingly common approach to this issue is to construct an asset-based index of SES. This is possible using surveys such as PIRLS, TIMSS and SACMEQ, where questions are asked regarding the ownership of certain household items. Filmer and Pritchett (2001) set forth a strong case that asset-based classifications of households correspond closely to classifications by expenditure, and that asset-based indices are in fact better at predicting educational attainment than expenditure information. This is an assuring finding from the point of view of investigating the effects of SES on education using PIRLS, TIMSS and SACMEQ, where there is no income or expenditure information.

The problem for asset-based indices is deciding how to derive an aggregate index from the range of asset variables in the data. One approach is to merely sum the assets in each household. However, this means that equal weight is given to the various assets, irrespective of how well they may predict SES. One can imagine that an asset owned by 98% of households would be less useful in differentiating SES than an asset that is owned by 50% of households. Equal weighting is therefore likely to be inappropriate for constructing an asset-based index of SES.

The solution to the question of weighting applied in this paper is to use Principal Components Analysis. This technique attaches the most weight to the asset variables that are most unequally distributed, i.e. the greater the standard deviation of the variable the greater the weight it is given in Principal Components Analysis (Vyas and Kumaranayake, 2006: 461). In Principal Components Analysis the range of variables is analysed so as to extract those linear combinations of the variables that capture the most common information (Filmer and Pritchett, 2001: 6). Each linear combination or “principal component” is uncorrelated with the others, so as to capture a different dimension in the data. The first principal component explains the most variation in the data with successive components explaining additional but less variation (Vyas and Kumaranayake, 2006: 460). Assuming a set of variables, X_1 to X_n, each principal component takes the following form:

$$PC_1 = w_{11} X_1 + w_{12} X_2 + \dots + w_{1n} X_n \qquad (1)$$

$$PC_m = w_{m1} X_1 + w_{m2} X_2 + \dots + w_{mn} X_n \qquad (2)$$

Where, w_{mn} is the weight for the *n*th variable within the *m*th principal component (adapted from Vyas and Kumaranayake, 2006: 460).

In this analysis only the first principal component (PC_1) is used for the construction of an index for SES. This is based on the critical assumption that the underlying concept explaining the linear combination with the most common information amongst the possessions variables is SES. Put differently, the SES of each student’s family causes the majority of the variation in the asset variables. The weights derived from the Principal Components Analysis are applied to the following formula for the overall asset index for each household *i*:

$$A_i = w_{11} * (a_{i1} - a_1)/(s_1) + \dots + w_{1n} * (a_{in} - a_n)/(s_n) \qquad (3)$$

Where w_{11} is the weight awarded to the first asset within the first principal component as determined in equation (1), a_{i1} is the value household *i* takes for asset 1, a_1 is the mean value of asset 1 for all households, and s_1 is the standard deviation for asset 1 over all households. (Adapted from Filmer and Pritchett, 2001: 6)

The student questionnaire in PIRLS 2006 captured information on whether students had access at their home to ten different items. The first six items were included in the questionnaires across all countries, and the last four items were country-specific. In the case of South Africa the ten items were the following: Computer, study desk/table, own books (excluding textbooks), newspaper, own room, own cellular phone, calculator, dictionary, electricity, tap water. This information provided ten variables suitable for use in Principal Components Analysis.[12] Therefore an SES index derived from Principal Components Analysis on the ten "possession" variables is used throughout the analysis to follow.[13] When comparing results from PIRLS with TIMSS and SACMEQ we use the same procedure to derive SES indices, making use of similar "possessions" questions that were administered in these surveys.[14]

[12] Another variable that is a good proxy for SES and is available in PIRLS is the educational attainment of parents. However, for various reasons all the analysis presented in the main text of this paper is based on an SES index using only the "possessions variables", and not parent education. Firstly, a decision needs to be made about whether to enter parent education into the Principal Components Analysis as a continuous variable, an ordinal categorical variable or as a set of binary dummies. This involves some fairly technical considerations. Secondly, regardless of which of these methods is used, the correlation coefficients between an SES index generated by including parent education and an SES index based on only the "possessions variables" are very high, at least in the case of South Africa (well above 0.9). Thirdly, and consequently, the results of the SES gradient analysis using an SES index that includes parent education are not substantively different to when the "possessions-only" index is used. These issues are more extensively discussed in Appendix C.

[13] Missing data is dealt with by imputation on the assumption that students who did not provide an answer did not have access to the relevant possession item. Three considerations informed this decision. Firstly, for the sake of sample size it seemed preferable to impute a value rather than drop observations. Secondly, one would intuitively expect an inability or unwillingness to give a definitive answer to be more common amongst students who do not have a particular item than amongst those that do possess the item. Thirdly, it was established that missingness (in the case of each possession variable) is negatively correlated with reading scores and is also negatively correlated with parent's education. This provides strong grounds for suspicion that failure to answer was most frequent amongst students of low SES.

[14] In TIMSS 2003 there were 16 "possessions" questions relating to the following items: calculator, computer, study desk/table, dictionary, electricity, running tap water, television, video player, CD player, radio, own bedroom, water flushed toilets, motor car, own bicycle, telephone, fridge. The possessions items in TIMSS 1995 and 1999 were similar but slightly different combinations. In TIMSS 1995 there were 16 items and in TIMSS 1999 there were 14 items. In SACMEQ there were 14 items: Daily newspaper, weekly/monthly magazine, radio, TV set, Video Cassette Recorder (VCR), cassette player, telephone, refrigerator/freezer, car, motorcycle, bicycle, piped water, electricity, table to write on. When we refer to SACMEQ, we are referring to SACMEQ II. SACMEQ I was carried out in 1993 and the exact same 14 items were included in the questionnaire, but South Africa did not participate.

3. Analysis of SES gradients

3.1 Generating SES gradients

An SES gradient is a graphical representation of the linear relationship between SES and a particular outcome of interest. This technique has commonly been applied in studies into the effects of SES on health outcomes. More recently, there has been a renewed interest in it's application to the effects of SES educational achievement (e.g. Willms,1997 & 2004, and Ross and Zuze, 2004). The procedure followed in this paper was to estimate a linear Ordinary Least Squares (OLS) regression with reading score as the dependent variable and the SES index, derived using the methodology explained in Section 2, as the explanatory variable.[15] One further adjustment was to standardise the SES index, converting it to have a mean of zero and a standard deviation of one. The equation therefore took the following form:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 SES_i$$

Where $\hat{Y}_i$ is the reading score, $\hat{\beta}_0$ is the intercept and $\hat{\beta}_1$ is the coefficient on the standardised SES index, which determines the slope of the gradient.

When running the regression for the South African data in PIRLS the following estimates were obtained[16] :

$$\hat{Y}_i = 301.44 + 59.29 * SES_i$$

This means that for every one standard deviation increase in student SES the predicted reading score was 59.29 points higher. The basic SES gradient for South Africa using these regression estimates is presented in Figure 5a.

---

[15] Ross and Zuze (2004: 8) note that there are negligible differences in the results produced by OLS and those by Hierarchical Linear Modeling (HLM).

[16] The coefficients were highly statistically significant. The R-squared value was 0.2223, indicating that 22.23% of the variation in reading scores was explained by the SES index.

Figure 5a: Basic SES gradient for South Africa

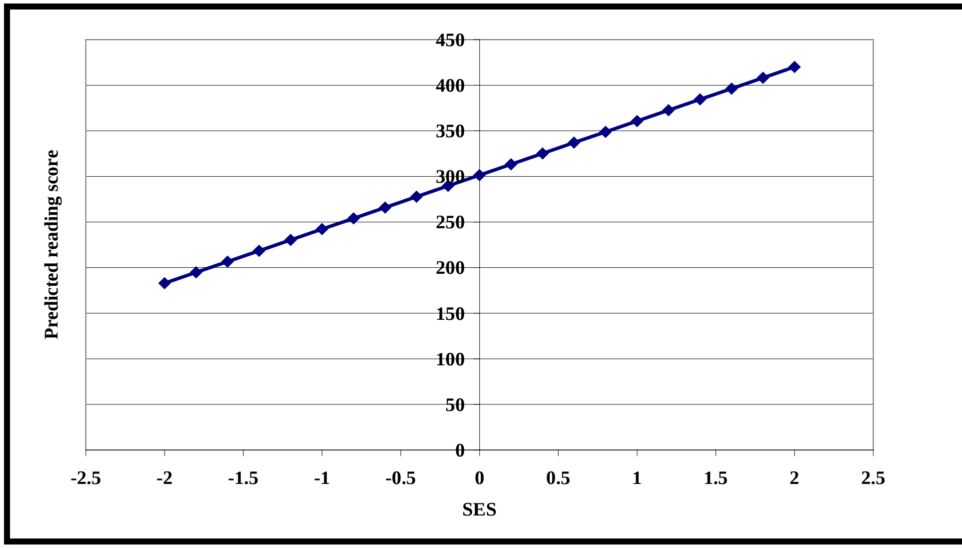


Figure 5b: SES gradient for SA with standardised reading scores

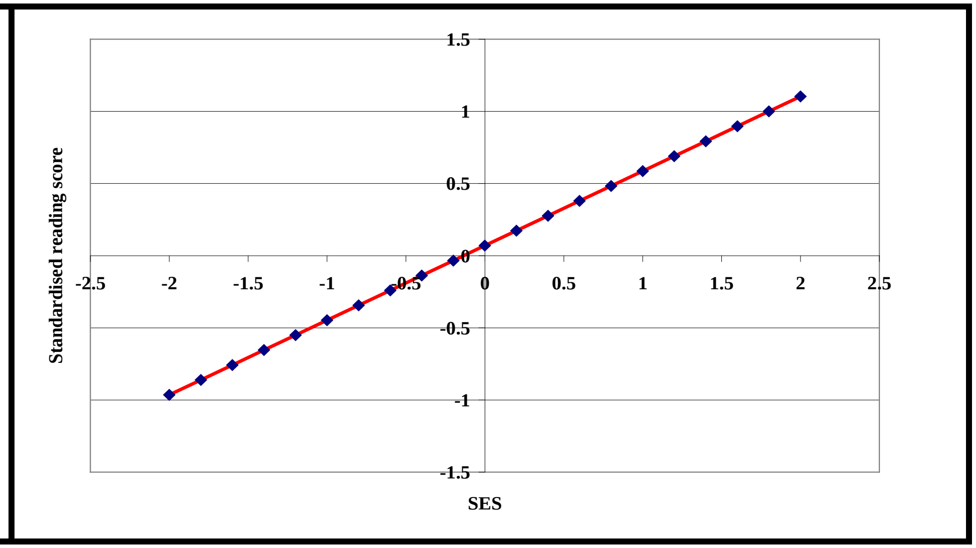


Figure 5c: SA SES gradients for reading surveys

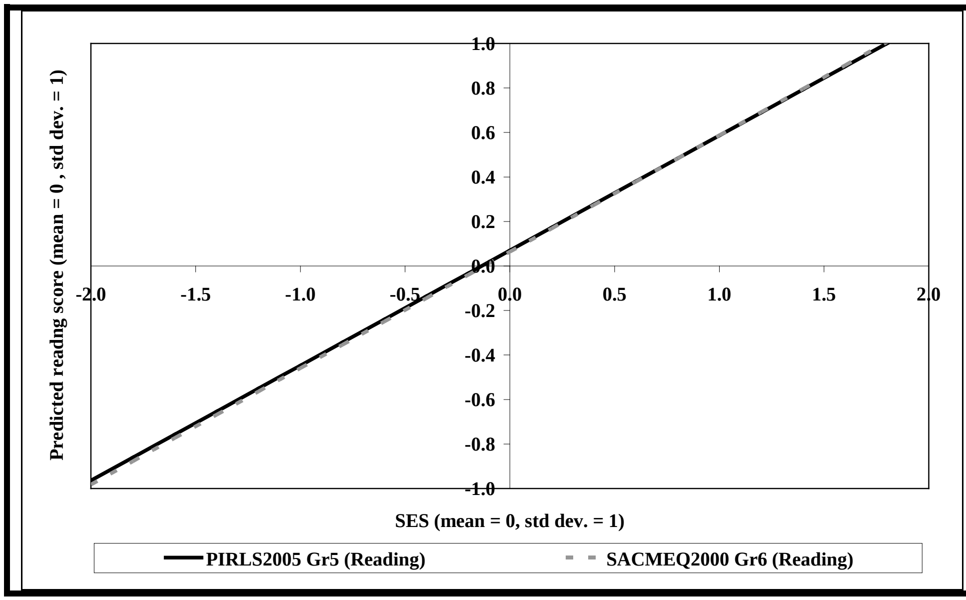
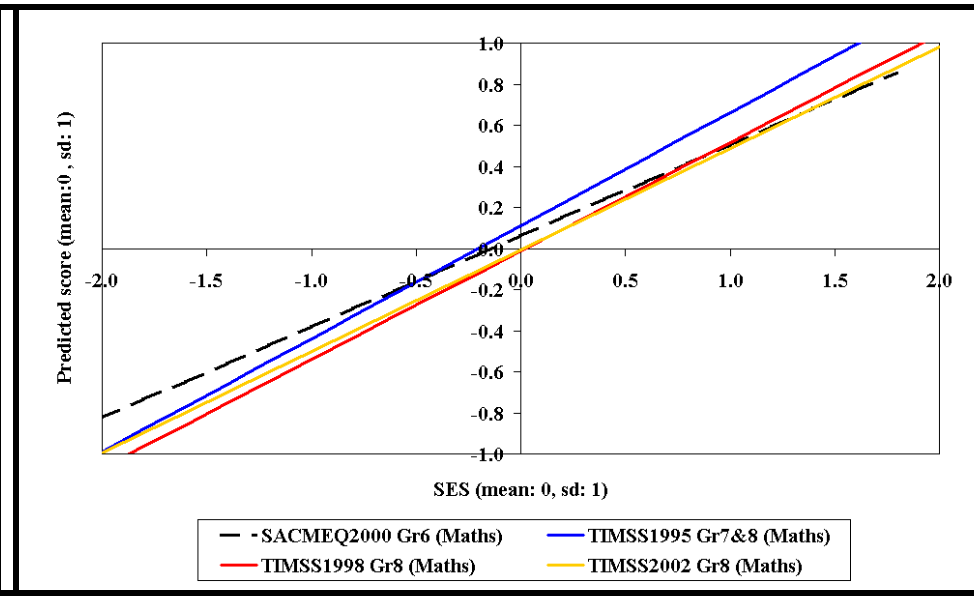


Figure 5d: SA SES gradients for Maths surveys



As Willms (2004: 7) explains, there are three components of the SES gradient that are important for the interpretation of these gradients such as that presented in Figure 5a. The *level* at the y-intercept is the expected reading score for a person with average SES. Thus for the basic SES gradient in Figure 5a, the person with average SES can be expected to have a reading score of 301.44. Note that this is not the mean reading score, although it happens to be very close. The *slope* gives an indication of the extent to which reading scores vary with SES. The *strength* refers to how much of the variance in reading scores can be attributed to SES. The R-squared statistic is commonly used as an indicator of the strength of the relationship.

For the sake of a more accessible interpretation the same equation was estimated except with standardised reading scores, i.e. with a mean converted to zero and standard deviation to one. The following estimates were obtained:

$$\hat{Y}_i = 0.0695 + 0.5169 * SES_i$$

Now the interpretation is that for every one standard deviation increase in student SES the predicted reading score increases by just over half a standard deviation. The basic SES gradient for South Africa with standardised reading scores on the Y-axis is presented in Figure 5b. Note that in this version of the SES gradient the level given by the intercept does not provide interesting information. We are now predominantly interested in the slope.

3.2 Comparison with SACMEQ and TIMSS

In order to gauge the reliability of our SES gradient we applied the same technique using data from SACMEQ II and the three waves of TIMSS (1995, 1999 & 2003). The SACMEQ data provides a useful link between PIRLS and TIMSS, in that SACMEQ tested reading (making it comparable with PIRLS) and mathematics (making it comparable with TIMSS). We constructed SES gradients for each survey and for each subject domain. The regression estimates are provided in Table 1 below:

Table 1: SES gradient regression estimates for PIRLS, SACMEQ & TIMSS[17]

Survey	Subject	Regression estimates
PIRLS	Reading	$\hat{Y} = 0.0695 + 0.5169 \times SES$
SACMEQ II	Reading	$\hat{Y} = 0.0637 + 0.5232 \times SES$
SACMEQ II	Mathematics	$\hat{Y} = 0.0630 + 0.4428 \times SES$
TIMSS 1995	Mathematics	$\hat{Y} = 0.1110 + 0.5502 \times SES$
TIMSS 1995	Science	$\hat{Y} = 0.1190 + 0.5726 \times SES$
TIMSS 1999	Mathematics	$\hat{Y} = -0.0121 + 0.5283 \times SES$
TIMSS 1999	Science	$\hat{Y} = -0.0241 + 0.5426 \times SES$
TIMSS 2003	Mathematics	$\hat{Y} = -0.0067 + 0.4948 \times SES$
TIMSS 2003	Science	$\hat{Y} = -0.0145 + 0.5085 \times SES$

Notice that the slope coefficients vary from 0.44 to 0.57. In order to gain a better perspective on how consistent the SES gradients are, consider Figure 5c and 5d, which presents the gradients according to subject. Figure 5c presents the gradients for the two reading tests (PIRLS and SACMEQ) and Figure 5d presents the gradients for the various mathematics tests (SACMEQ and TIMSS). The gradients for the science tests are shown in Appendix E.

What is perhaps most pleasing for the sake of this paper is how similar the SES gradients are for the two reading surveys. At first glance it looks like there is only one gradient presented in Figure 5c. However, a closer look shows that the gradient for the SACMEQ reading scores lies virtually on top of our original SES gradient for the PIRLS data. The slope coefficients are very close: 0.5169 for PIRLS and 0.5232 for SACMEQ.

Turning now to the mathematics surveys in Figure 5d, the slopes for the various SES gradients are fairly consistent across the three waves of TIMSS. The gradient for the SACMEQ data is somewhat flatter however. One possible contributing factor to this result relates to the fact that

[17] Appendix D shows more detailed information on TIMSS and SACMEQ as well as details on how the respective SES indices were derived. Moreover, estimates of SES gradients using a quadratic function (including SES-squared) are presented.

SACMEQ tested at grade 6 level whereas TIMSS tested grade 8 students. Given that much of the South African school system is seriously under-performing (as Section 2 demonstrated), and that much of the variance in educational achievement is explained by SES, one might expect low-SES students to fall further behind with every grade they move through. One would therefore expect low-SES students to be further behind high-SES students by grade 8 than in grade 6. This would manifest in a steeper SES gradient at the grade 8 level than at the grade 6 level, which is what we do see in Figure 5d.

The comparison with TIMSS and SACMEQ is encouraging from the point of view of the methodology employed in this paper, as it would appear that the results are largely consistent across different datasets. From hereon the focus returns to the PIRLS data.

3.3 International Comparison

SES gradients were constructed for three other participants in PIRLS in order to get some sense of how important SES is as a determinant of educational outcomes in South Africa by international standards. Russia was selected because it was the top performing country in PIRLS 2006. Morocco was selected because it is most similar to South Africa in that is the only other African participant and in that it achieved the second-lowest average reading score (323). The USA was chosen as it is fairly well established that educational performance varies strongly with SES in the USA (e.g. Willms, 2004).

Figure 6a presents the SES gradients for South Africa, Russia, Morocco and the USA, while Figure 6b presents the gradients with the standardised reading score. Table 2 shows the various estimates and regression statistics. It is not surprising that the level of the SES gradients is far lower for South Africa and Morocco than for Russia and the USA. This reflects the vast gap in overall reading performance. The slopes of the gradients provide information on how different the relationship between reading scores and SES is across these countries. It is striking how much steeper the South African SES gradient is than all the others. The standardised version in

Figure 6a: International Comparison of SES gradients

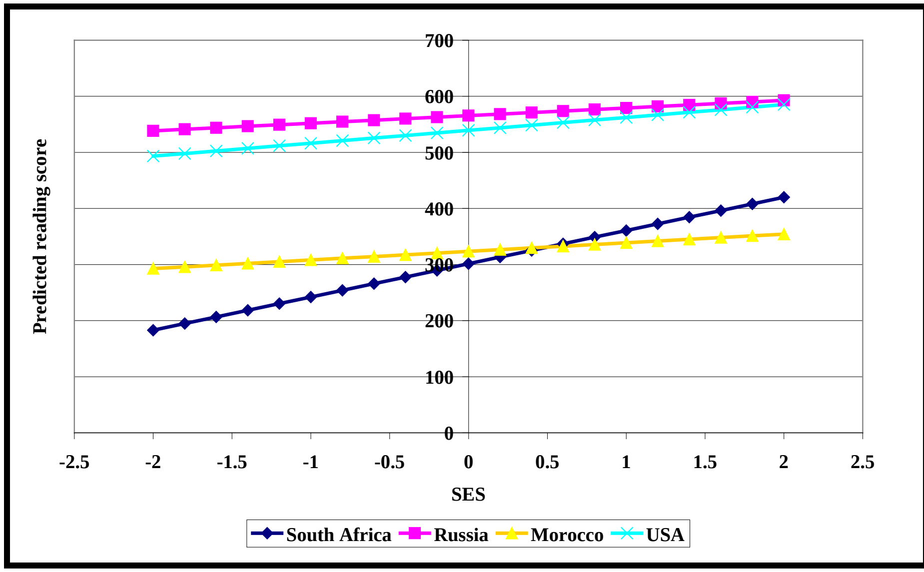


Figure 6b: Country gradients using standardised reading scores

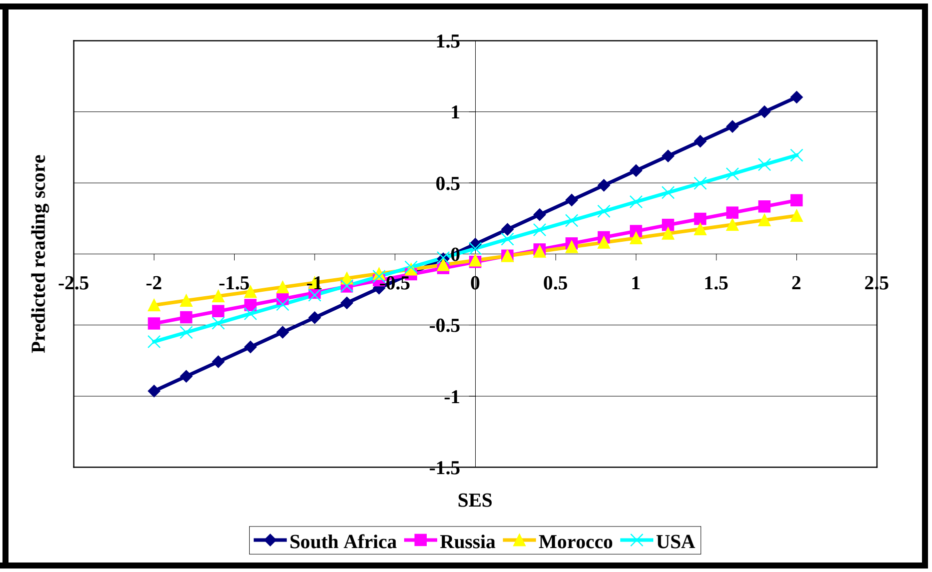


Table 2: Estimates and regression statistics for SES gradients by country

	South Africa	Russia	Morocco	USA
Intercept*	301.44	565.58	323.52	539.33
Coefficient on SES*	59.29	13.60	15.36	22.97
Coefficient on SES**	0.52	0.22	0.16	0.33
t-statistic	64.73	14.8	8.79	25.07
p-value	0	0	0	0
R-squared	0.2223	0.0444	0.0232	0.1081

*Using reading score as the dependent variable
**Using standardised reading score as the dependent variable

Note that the t-statistics, p-values and R-squared statistics are the same irrespective of whether the dependent variable is standardized or not.

Figure 6b makes it easier to see how different the slopes are. Comparing South Africa and Morocco, we see that although these countries both have low average reading scores, these scores vary far more with SES in South Africa (with a slope coefficient of 0.52) than in Morocco (with a slope coefficient of 0.16). The South African SES gradient is even considerably steeper than that of the USA, where it is well known that educational outcomes vary substantially with SES differences. Thus we can conclude that in South Africa there are very large differences in reading scores across socio-economic groups, by international standards.

The strength of the relationship between SES and reading scores is also very different amongst the four countries compared in Figure 6. Consider the R-squared values in Table 2. The values are low for Russia and Morocco, indicating that very little of the variance in reading scores is attributable to differences in SES. However, for the USA and for South Africa high R-squared values are obtained. This result for the USA is consistent with that of Willms (2004: 8), who found that the strength of the SES gradient for the USA was significantly greater than that for Canada. It is therefore concerning that we find the relationship between SES and reading scores to be more than twice as strong in South Africa as in the USA – the R-squared for South Africa is 0.2223 and for the USA is 0.1081. One way of interpreting the R-squared value is as a measure of how deterministic SES is for educational achievement. Thus we can say that in the case of South Africa a student with a given SES has more than twice the chance of achieving a reading score approximately equal to the reading score predicted by the SES gradient, than would be the case in the USA.

This international comparison of SES gradients adds to what we know about the generally dismal performance of the South African school system. We see that there is wide inequality of reading achievement across different socio-economic groups, and that this relationship seems to be more deterministic than in other countries. Taken together, these findings raise concerns regarding the prospects for social mobility in South Africa. This question will be addressed in Section 5.

3.4 School level analysis of SES gradients

The analysis is now extended by constructing SES gradients for South Africa, Russia, Morocco and the USA with schools as the units of analysis. In Section 1 it was theorised that SES may affect educational achievement through various aspects of context such as home support, neighbourhood effects, within-school peer effects, school-level resources, etc. This motivates the construction of SES gradients at the school level of analysis on order to capture more of these effects, which tend to be concentrated amongst groups of similar SES.

This is achieved by plotting the linear equation of the estimates from a regression of reading scores on the school mean SES[18]. Note that the dependent variable is not each school's mean reading score, but is still each student's reading score. This makes the slopes and strength of the gradients directly comparable with the analysis presented in the preceding section. Our earlier gradients showed the effect of a student's own SES on her reading score. Now, the SES gradients at the school level show the effect of the SES of the school that a student attends on that student's reading score. The results are presented in Figure 7. As before, Figure 7a presents the SES gradient with reading score (not standardised) as the dependent variable, thus capturing the slope of the gradients as well as the level. Figure 7b presents the gradients with the standardised version of the reading scores in order to highlight the differences in the respective slopes. Table 3 shows the various estimates and regression statistics.

The results shown in Figure 7 make for some interesting comparison with the earlier SES gradients presented in Figure 6. As before, all the regressions are highly significant. Once again South Africa has comfortably the steepest gradient. The most consistent difference between the student-level SES gradients and the school-level SES gradients is that all the slope coefficients and all the R-squared values are higher for the school-level gradients.

The most substantial increases occur in the cases of South Africa and Russia.[19] A one standard deviation increase in the SES of a Russian student leads to an increase in that student's predicted reading score of 13.60 points. However, a one standard deviation increase in the mean SES of the school a student attends leads to an increase in that student's predicted reading score of 24.05 points. The corresponding slope coefficient in the standardised version of the gradient increases from 0.22 to 0.38. Moreover, the R-squared value for Russia's school-level SES gradient is dramatically higher than for the student-level gradient – 0.15 as opposed to 0.04 – indicating that

[18] This is simply the mean of the SES scores for all the students in a school, using our SES index as derived from the ten "possessions" questions.
[19] The slope coefficients and R-squared values also increase for Morocco and the USA, although these increases are less substantial than in the cases of Russia and South Africa and will therefore not be discussed at length.

Figure 7a: School SES gradients in International Comparison

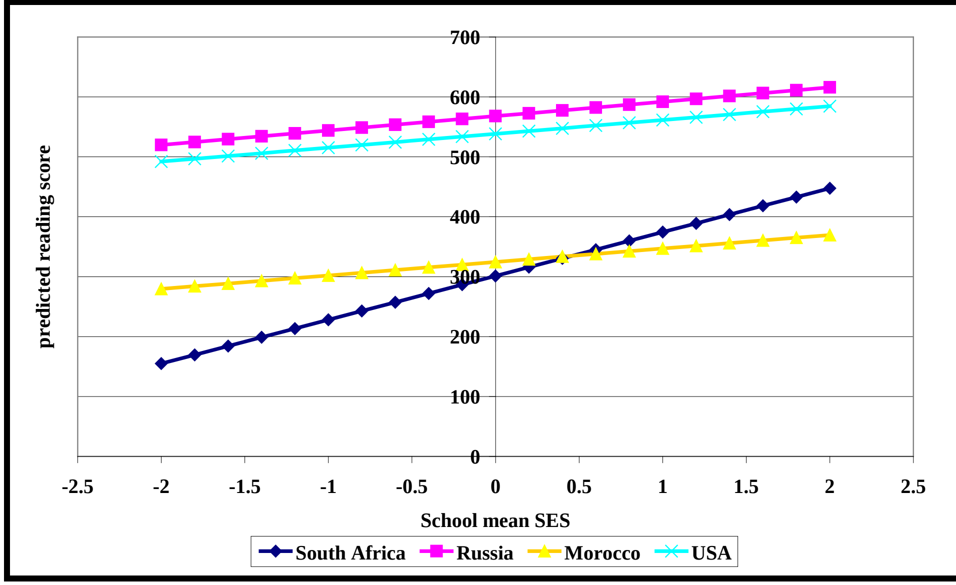


Figure 7b: School gradients with standardised reading scores

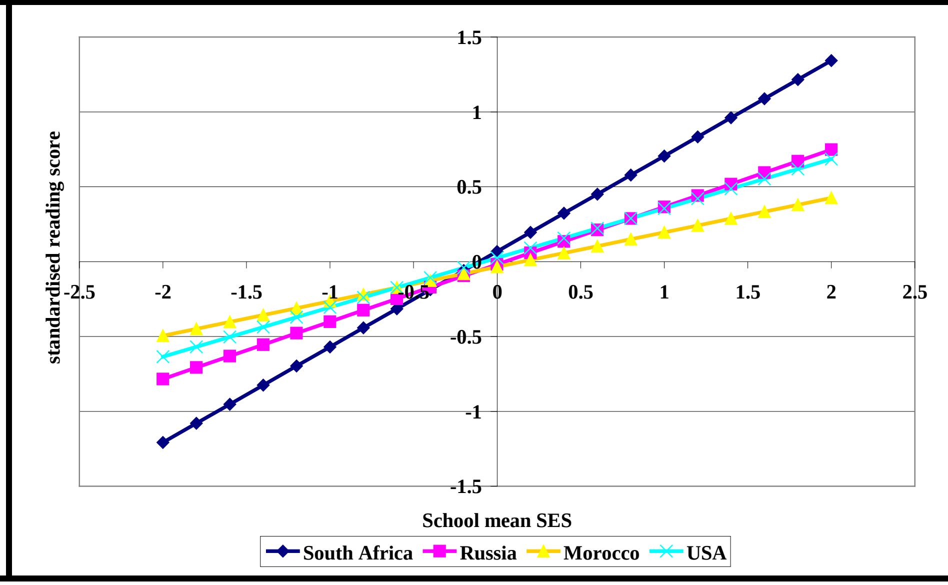


Table 3: Estimates and regression statistics for School SES gradients by country

	South Africa	Russia	Morocco	USA
Intercept	301.25	567.95	324.53	538.38
Coefficient on SES*	73.14	24.05	22.49	23.11
Coefficient on SES**	0.64	0.38	0.23	0.33
t-statistic	103.15	28.85	13.46	26
p-value	0	0	0	0
R-squared	0.42	0.15	0.05	0.12

*Using reading score as the dependent variable
**Using standardised reading score as the dependent variable

Note that the t-statistics, p-values and R-squared statistics are the same irrespective of whether the dependent variable is standardized or not.

far more of the variance in student reading scores is explained by school SES than by student SES.

In the South African case, the student-level analysis in the previous section showed that a one standard deviation increase in a student's SES leads to an increase in that student's predicted reading score of 59.29 points. However, when the mean SES of a student's school increases by one standard deviation the predicted reading score for that student increases by 73.14 points. The corresponding slope coefficient in the standardised version of the gradient increases from 0.22 to 0.38. The R-squared value is also substantially higher in the school-level SES gradients than in the student-level gradients – 0.42 as opposed to 0.22 – indicative of the greater strength of the school-level gradient. Put differently, far more of the variance in student reading scores in South Africa is explained by school SES than by student SES.

Another way of gaining some perspective on the relative contributions to reading scores of school factors *vis a vis* student factors is by generating intra-class correlation coefficients. The intra-class correlation coefficient (rho) is given by the following formula:

$$\rho = V(\mu_{0j})/[V(\mu_{0j}) + V(r_{ij})]$$

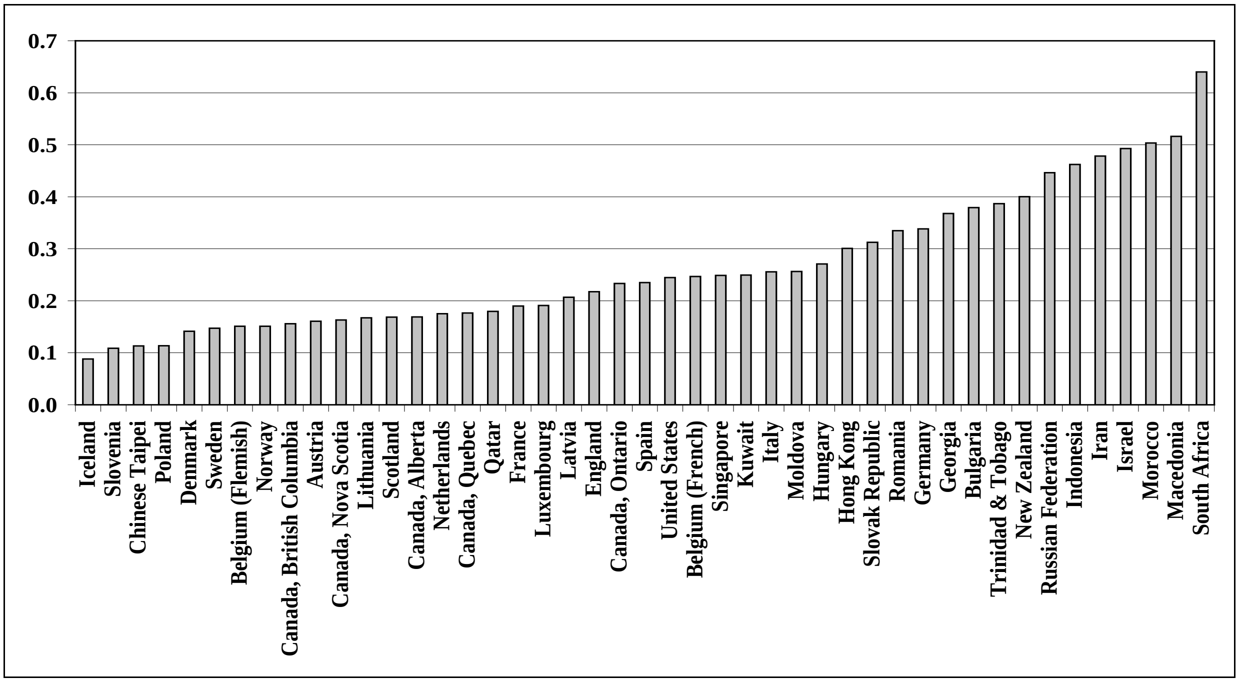
Where $V(\mu_{0j})$ is the between school variance and $V(r_{ij})$ is the within-school variance.[20] According to the formula an index is created between zero and one. This rho value captures the proportion of the overall variance in reading scores explained by between-school differences as opposed to within-school differences amongst individuals. Therefore, in a country where each school was identical 'rho' would equal zero. Conversely, if all the variation in reading scores was determined by school-level factors the rho value would be equal to one.

We generated rho values for all the participants in PIRLS 2006. Note that the rho values say nothing about the impact of SES or any other student or school-level characteristic on reading scores. They only indicate how much of the overall variance in reading scores can be attributed

[20] $V(\mu_{0j})$ and $V(r_{ij})$ are derived using a fully unconditional Hierarchical Linear Model. This technique is explained in Appendix F.

to within-school variance and how much to between-school variance. The results are presented in Figure 8.

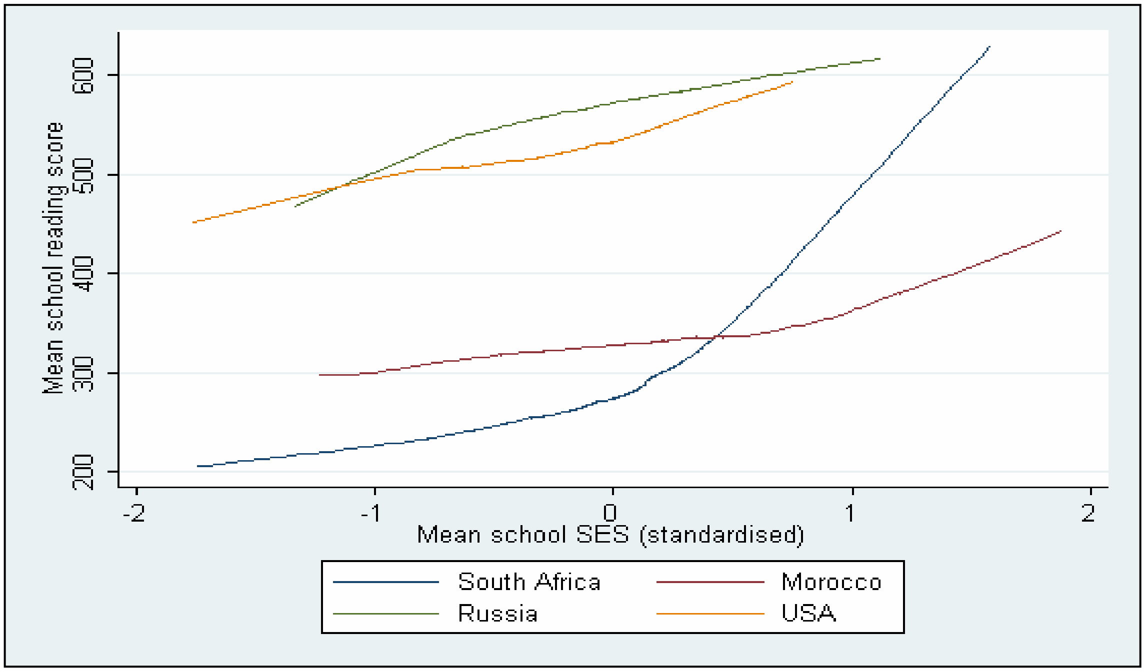
Figure 8: Intra-class correlation coefficients (Rho)



It is striking that South Africa has the highest rho value by a considerable margin, indicating that the proportion of the overall variance in reading scores that can be attributed to between-school differences is greater in South Africa than in any of the other countries. This would imply that a critical dimension to understanding the overall performance of South African students in PIRLS 2006 is to focus on differences in school quality throughout the system. Of course the rho values do not necessarily imply that SES is an important determinant of school quality, but the school-level SES gradients provide enough evidence to motivate a further investigation.

Before moving on to multivariate analysis, one more type of socio-economic gradient is presented in Figure 9 for the sake another perspective of the relationship between SES and reading achievement. The Lowess regressions for South Africa, Russia, Morocco and the USA differ from the gradients presented thus far in two main respects. Firstly, a Lowess regression does not require a linear or quadratic model specification but carries out locally weighted regressions at each data point and smoothes the result through the weighting system. This means that the shape of the Lowess curve is determined by the data rather than by the imposition of a model specification. Secondly, the dependent variable is now school mean reading score instead of student reading score, as was the case in all of the above SES gradients.

Figure 9: Lowess regressions in international comparison



Various observations should be made from Figure 9. As before, the level of the gradients is substantially higher for Russia and the USA than for South Africa and Morocco, due to the high overall performance of the former countries in PIRLS. More interestingly though, it appears that the relationship between SES and reading performance is approximate to a linear relationship in the cases of Russia, Morocco and the USA, whereas in the case of South Africa the relationship appears to be non-linear and convex. This means that SES seems to be more strongly correlated with reading achievement at higher levels of SES. This observation is returned to and investigated in sections 3 and 4 of this paper. It certainly warrants the inclusion of the square of SES and the square of school mean SES in the multivariate analysis to follow.

To summarise this section, it is evident that in South Africa, SES plays an exceptionally strong role in determining educational achievement. The effects of SES appear to be intensified through schools. It seems that the combined SES of the school a particular student attends may be more important than her own family background. More light on this issue is shed by the multivariate analysis reported in the following section.

4. Multivariate analysis on South African student achievement in reading

In this section we explore the impact of a range of variables provided in PIRLS on reading scores. In this way we can further investigate the role of SES by controlling for a variety of other factors that influence reading achievement. The analysis in this section is limited to South Africa.

The regressions in this section all have the standardised version of students reading scores as the dependent variable. Thus the reading scores are converted to have a mean of zero and standard deviation of one. In total 26 survey regressions were run.[21] The results of these are presented in Appendix H. In the main text (Table 4) the results of the most important eight regressions are presented. Four categories of explanatory variables were included – variables relating to student characteristics (student-level), variables relating to the home environment (home-level), variables relating to teacher characteristics (teacher-level) and those relating to school-specific characteristics (school-level).[22]

The first student-level variable is SES as defined and used in the construction of SES gradients in Section 3. Note that the standardised version of SES is used (zero mean and standard deviation of one). The other student-level variables include the squared SES, age dummies for whether students are above or below the average age in grade 5, a gender dummy, provincial dummies, dummies for the frequency of being given homework, dummies for the frequency of speaking the language of the test at home, dummies for how often students receive help with work at home, dummies derived from an index of student's attitude toward reading, dummies derived from an index of how safe students feel at school, dummies for how often students use books from the school or local library and dummies for the number of books a student has at home.

The home-level variables include a dummy for pre-school attendance, dummies based on an index for home-based reading activities before school and dummies for parent's education. The teacher-level variables include class size, dummies for the teacher's highest educational

[21] The survey design variables were identified as follows: The stratum variable is province (of which there are 9), and the Primary Sampling Unit (PSU) variable is schools (of which there are 397).
[22] These categories of variables correspond to the various questionnaires administered in PIRLS, i.e. the student questionnaire, the home questionnaire, the teacher questionnaire and the school questionnaire.

attainment and the percentage of time spent on teaching as opposed to other activities such as discipline. The school-level variables include the language the reading test was administered in, dummies for whether the school was situated in a rural, suburban or urban area, an index for school resource shortages, dummies to capture the severity of the problem of absenteeism, dummies for the availability of a library and how many books it has, the school mean SES and the square of school mean SES. Lastly a variable for the proportion of students in each school scoring at or above the national average (302) was included.[23]

Table 4 shows the results of the regression analysis. Regression [1] is the basic SES gradient equation with only SES as an explanatory variable. Regression [2] is the quadratic version of the SES gradient with SES-squared included. Regression [3] adds all the student-level variables. Regression [4] includes the home-level variables. In regression [5] the teacher-level variables are added. Regression [6] introduces the school-level variables, but excludes school mean SES, which is added in regression [7]. Finally, regression [8] includes variable for the proportion of students in each school scoring at or above the national average (302).

[23] The full range of the variables is described in Appendix G.

Table 4: Regressions on the standardised reading score

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Student-level variables								
SES	0.5169**	0.5597**	0.3400**	0.3115**	0.3020**	0.1750**	0.0693**	0.0687*
SES-squared		0.2166**	0.1757**	0.1589**	0.1307**	0.0795**	0.0292**	0.0185
Under11			-0.3751**	-0.3357**	-0.3090**	-0.2150**	-0.2033**	-0.1788**
Over11			-0.4595**	-0.4649**	-0.4708**	-0.3268**	-0.3151**	-0.1991**
Female			0.1615**	0.1581**	0.1605**	0.1915**	0.1964**	0.1987**
Speak1			0.2614**	0.2464**	0.2396**	0.1923**	0.1505**	0.1210**
Homework1			0.1401**	0.1368**	0.1241**	0.1215**	0.1271**	0.1237**
Homework2			0.2366**	0.2338**	0.1951**	0.1149*	0.1089*	0.0838*
NoHelp			0.3291**	0.3045**	0.2909**	0.1837**	0.1533**	0.1397**
HelpParent			0.2381**	0.2043**	0.1858**	0.1175**	0.0879**	0.0584*
Attitude1			0.3457**	0.3277**	0.3248**	0.2891**	0.2971**	0.2634**
Safety1			0.1655**	0.1587**	0.1640**	0.1645**	0.1466**	0.0992**
Library1			0.1149**	0.1102**	0.1133**	0.0557	0.0566*	0.0444*
Book11			0.1995**	0.1764**	0.1476**	0.0741*	0.0355	0.0379*
Home-level variables								
PreSchool				0.0352	0.0460	0.0454	0.0087	0.0199
Early1				0.0734**	0.0671**	0.0895**	0.0706**	0.0691**
ParentMatric				0.3362**	0.3169**	0.2106**	0.1579**	0.1327**
Teacher-Level Variables								
ClassSize					-0.0096**	-0.0062**	-0.0034*	0.0001
TPostMatric					0.1684**	0.0704	0.0597	-0.0136
TTeaching%					0.0073**	0.0012	0.0002	0.0003
School-Level Variables								
Afrikaans						0.7267**	0.5934**	0.1845**
English						0.3919**	0.1604	-0.0361
UrbanSub						0.1931*	0.0254	-0.0729
SResource						-0.0179	-0.0049	-0.0136*
SAbsent1						-0.2494**	-0.1204	0.0147
SAbsent2						-0.1366*	-0.0612	-0.0096
SLibrary1						0.8995**	0.4482**	0.2102*
SLibrary2						0.2371*	0.0694	-0.0415
SSES							0.3623**	0.1687**
SSES_squared							0.1314**	0.1048**
Pass%								1.6084**
Constant	0.0695	-0.1567**	-0.9032**	-1.0179**	-1.0256**	-0.7918**	-0.7507**	-1.1961**
Sample size	14 657	14 657	14 343	14 343	11 942	10245	10245	10185
R-squared	0.2223	0.2647	0.4828	0.5042	0.5332	0.6324	0.6723	0.7276

* Significant at 5%, ** Significant at 1%

Note: The above regressions control for province. The full results are reported in Appendix H.

.

The coefficients on SES were positive and significant across all the regressions. However, they became smaller as the other explanatory variables were added. It is noteworthy that according to regression [6], for example, a one standard deviation increase in SES is associated with 0.175 of a standard deviation increase in reading score, once all the student, home, teacher and school level variables have been controlled for. The coefficients on SES-squared were positive and significant at the one percent level for all but regression [8]. This would suggest that the relationship between SES and reading achievement is non-linear and convex. Put differently, there appears to be a stronger association between reading scores and SES at higher levels of SES. This is consistent with the shape of the Lowess regression for South Africa as presented in Figure 9.

Turning to the student-level variables, it can be seen that the two age dummies were negative and significant at the one percent level in all regressions, indicating that students who are either too young or too old for grade 5 are at a disadvantage and tend to have lower reading achievement. The female dummy was positive (between 0.16 and 2) and significant in all regressions. The provinces that performed better than the Eastern Cape, which is the reference (or omitted) province, had positive and significant coefficients in regression [3]. However, as more teacher and school variables were added the province dummies lost their significance, with the exception of KwaZulu Natal, which maintained its positive and significant coefficient. Students that spoke the language of the test at home either sometimes or always, performed better according to all the regressions. The more frequently homework was given the better students performed. Students that either do not need help ('nohelp') or that do receive help from a parent or grandparent perform better than students who have no help available to them, except perhaps from a sibling. This result was significant across all regressions. There is reason to believe that this variable relates to an important dimension of SES in South Africa – having no help available may be indicative of disruptive family and household structures such as skip-generation households, a feature which is known to be common amongst poor households in South Africa. Family and household structure may well be one important transmission mechanism between SES and reading achievement in South Africa. The two dummy variables derived from the two student indices – attitude towards reading, and safety at school – were positive and significant. If students borrowed books from the school or local library at least once a week or had more than ten books at home, this was associated with higher reading scores. Of course these variables can be hypothesized to influence reading ability directly, although they may well be indicators of the

SES of the school (in the case of the library variable) and of the SES of the student (in the case of the 'number of books at home' variable).

Turning to the home-level variables, we see that the coefficient on having attended pre-school for at least three years was positive and significant at the five percent level in only some regressions (In regressions [K], [R], [S],[T] and [V] in Appendix H). The coefficient on the index for early childhood reading-based activities was consistently positive and significant. If either parent had at least matric education this was strongly associated with higher student reading achievement. The coefficient was consistently positive, large and significant at the one percent level. These home-level variables are relevant to the investigation into the effect of SES on reading achievement, as parent's education (which also feeds into early childhood reading activities) is an important proxy for SES. Indeed, it was a strong candidate for inclusion in the index for SES, but was left out for technical reasons, as explained earlier.

The teacher-level variables were all significant when included in regression [5]. Bigger class sizes had a small but significant negative effect on reading scores. There was a positive relationship between the education of teachers beyond matric and student reading scores. Similarly, the higher the proportion of time spent on teaching as opposed to other activities, the better the associated reading scores. The remaining teacher-level variables described in Appendix G – the index for the variety of activities undertaken in class, the experience, age and gender of teachers, and the index for teacher career satisfaction – were found to be insignificant and were thus excluded from the regression analysis. The variables for the education of teachers beyond matric and the proportion of time spent on teaching as opposed to other activities lost their significance once the school-level variables were included. Of the teacher-level variables only class size remained significant once school-level variables were controlled for, although the size of the coefficient on class size was very small. That the other teacher-level variables were not statistically significant is unsurprising given that it is well known that better quality teachers and teachers with higher levels of education tend to be concentrated in the schools with higher levels of SES. Moreover, better-resourced and higher SES schools generally provide an environment more conducive to higher teacher effort.

The inclusion of the school-level variables substantially increased the explained variance in reading scores, as indicated by the R-squared value, which increased from 0.53 in regression [5] to 0.63 in regression [6] and 0.67 in regression [7]. As regression [6] shows, schools that took the test in English or Afrikaans achieved higher reading scores than schools that took the test in one of the African languages. This language variable probably roughly divides schools according to former education department[24] – where the group that took the test in an African language probably comprises only historically black (former DET) schools and the group that took the test in English or Afrikaans probably includes all the historically white, coloured and Indian schools as well as some historically black schools. Once school mean SES was included in regression [7] and [8] the 'english' dummy was no longer significant. This is probably because school SES now captured much of the effect of former education department. The "Afrikaans" dummy remained positive and significant throughout and this is understandable given that a fairly wide variety of schools would have written in English but it is unlikely that many historically black schools would have written in Afrikaans.

Students in urban or suburban areas tended to perform better than those in rural areas, although this effect was not significant once the school SES was controlled for. Shortages in school resources were associated with lower reading scores, although this result was only significant in some regressions ([T], [U] and [V] in Appendix H). The more seriously schools reported struggling with the problem of absenteeism, the lower the reading performance. However, once mean school SES rather than just individual SES was controlled for the significance of the school absenteeism dummies was lost. This indicates that school absenteeism is predominantly a problem amongst poorer schools. This is one indicator of the various problems of inefficiency that schools with low levels of SES suffer from in South Africa. If a student attended a school with a well-stocked library this was associated with a higher predicted reading score, although this effect was less pronounced once mean school SES was controlled for.

The inclusion of mean school SES ('SSES') and it's square ('SSES_squared') increased the explained variance of reading scores appreciably and resulted in many of the teacher and school-level variables becoming insignificant. Both 'SSES' and 'SSES_squared' had positive and significant coefficients, which were relatively large. The fact that the coefficient on 'SSES_squared' was positive, large and significant provides further evidence of a convex

[24] Under the apartheid regime there were separate education departments according to race. The Department of Education and Training (DET) was the department for schools with black students.

relationship between SES and educational achievement. It would appear that improvements in SES only begin to be associated with considerably better reading scores at higher levels of SES.

The inclusion of ‘SSES’ and ‘SSES_squared’ in regressions [7] and [8] resulted in noticeably smaller coefficients on student SES and the square of student SES. In regression [7], the coefficients on both SSES (0.3623) and SSES_squared (0.1314) were considerably larger than the coefficients on student SES (0.0693) and the square of student SES (0.0292), indicating that the combined SES of a school has a much bigger impact on a student’s reading score than that student’s own SES. There are various channels through which one might expect the SES of a school to influence reading achievement. These include institutional factors such as organizational inefficiencies, school problems, absenteeism, resource shortages and poor quality resources (such as poor teacher quality). The regression analysis has already provided some preliminary evidence of this in that many of the teacher and school-level variables were significant before school SES was included in the model specification, but not after its inclusion. Another channel through which school SES might be expected to influence student outcomes is through peer effects. This was one of the major findings of the Coleman Report, as discussed in Section 1. One might expect that when a group of students with low levels of home educational support, parental expectations and personal life aspirations are put together in school, the effects of these factors will be compounded. This is not to mention an array of neighbourhood factors (such as gangsterism, crime, alcoholism amongst the men of the community and a pervasive atmosphere of hopelessness) that is not conducive to learning and that persists in many poverty-stricken areas of South Africa.

Another peer effect is the general culture of learning in a school and the performance of other students. In an attempt to capture some of this, the proportion of students scoring above the national average in each school was included as an explanatory variable in regression [8]. The coefficient on “Pass%” was positive and significant at the one percent level. Of course the coefficients on all the other explanatory variables became smaller and many lose their significance. However, it is noteworthy that when controlling for all these factors, the performance of classmates had an important impact on reading achievement. This result may incorporate some unobserved aspects of school quality, but it may also say something about a peer effect.

A final observation from the regression analysis is that the R-squared value of 0.7276 in regression [8] is quite high for this kind of regression, indicating that about 73% of the variation in reading scores is explained by this model. This is noteworthy considering that something as important as innate ability is not included in the model.

The multivariate analysis has confirmed that SES has an important impact on educational achievement in South Africa. Moreover, it has emphasized the major role that schools play as the settings where the effects of SES are compounded. The next section investigates this somewhat further and considers the implications of the major influence of SES on educational achievement on the prospects for social mobility in South Africa.

5. The prospects for social mobility

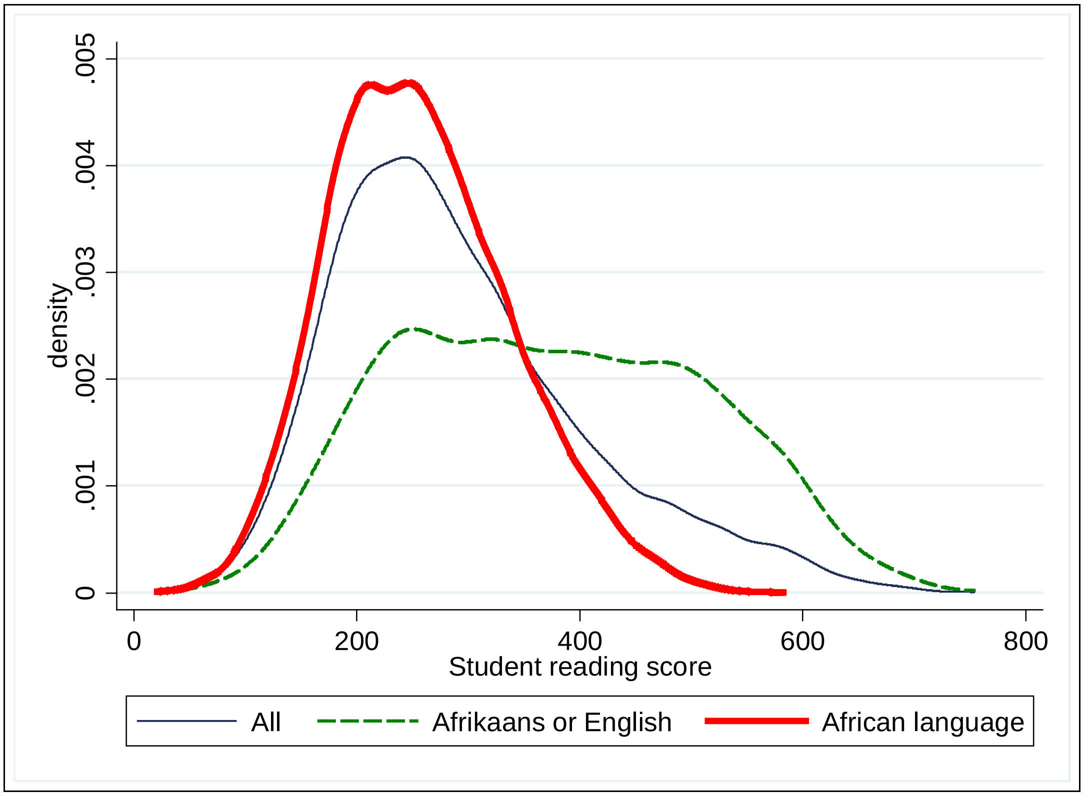
5.1 The nature of educational inequality in South Africa

The previous sections have demonstrated that educational achievement in South Africa varies considerably with differences in SES. It also well known that there is a strong racial dimension to the distribution of SES in South Africa, attributable to the country's long history of institutionalised inequality on the basis of race. Moreover, the country has historically had separate education systems for the respective race groups. In this section a rather crude attempt is made to distinguish between the historically disadvantaged and predominantly black schools and the historically affluent and predominantly White, Coloured and Indian schools. The aim is to explore how this distinction sheds light on the apparent relationship between SES and educational achievement. The section ends with a consideration of the prospects for social mobility in South Africa given what has emerged about the distribution of educational achievement in the country.

PIRLS does not offer any direct information on race. The best available indicator is the language that each school chose to take the test in. This permits the assumption that if a school took the test in one of the African languages then the school probably belongs to the historically disadvantaged and predominantly black school system. Dividing the PIRLS data for South Africa on this basis into historically black schools and historically white, coloured and Indian schools results in about 70% of the observations falling into the historically black school system. In reality about 80-85% of South Africa's schools might be expected to fall into this category. Therefore it is likely that this division on the basis of language underestimates the number of historically black schools. This is because some historically black schools may have taken the test in English, and therefore been misclassified. However, it is highly unlikely that the opposite error has been committed – there is no reason to expect any predominantly white, coloured or Indian schools to have written the test in a language other than English or Afrikaans.

Figure 10 below presents kernel density curves of the reading scores by the language of the test. It is clear that two very different distributions arise. The distribution of reading scores for the historically black schools lies to the left of the Afrikaans/English distribution and is much more concentrated, whereas the Afrikaans/English reading scores have a greater variance.

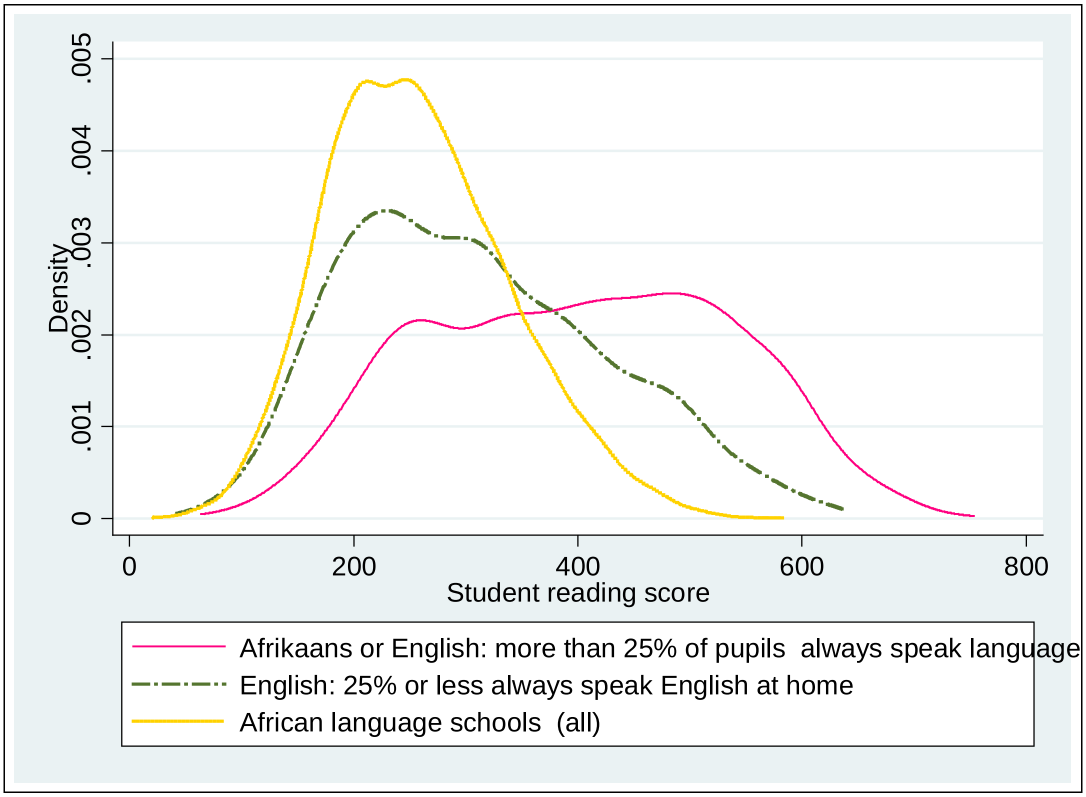
Figure 10: Kernel density curves of student reading scores[25] by language of test



As mentioned above, there are likely to be a number of predominantly black schools that took the test in English. Such cases will have been misclassified into the "Afrikaans or English" category. One way to attempt to identify such schools is to look for schools that took the test in English but where the majority of the students do not usually speak English at home. In Figure 11 kernel density curves of the reading scores by the language of the test are presented, including a separate category for schools where the language of the test was English, but where more than 75% of the students do not "always" speak English at home.

[25] In this kernel density curve the reading scores are standardised to have a mean of zero and standard deviation of one.

Figure 11: Kernel Density Curves by language of test and frequency spoken

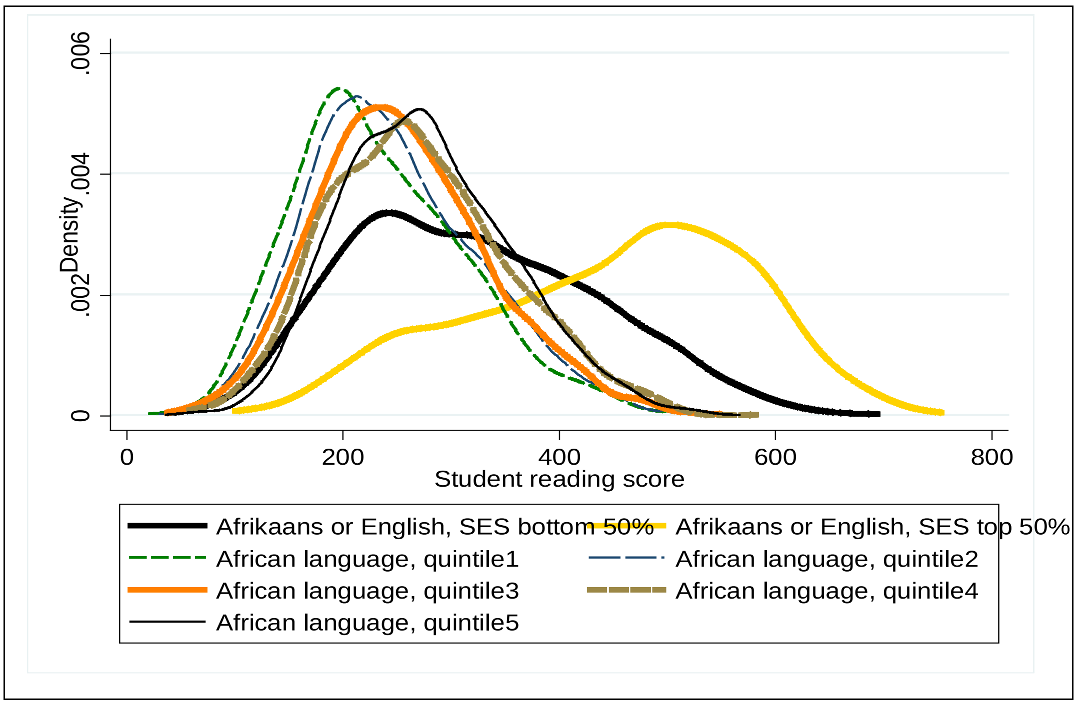


This distinction produces a distribution for the supposedly privileged schools that is more skewed to the right than in Figure 10. The distribution for the group of schools where the language of the test was English, but where more than 75% of the students do not “always” speak English at home looks somewhat similar to the distribution of reading scores for the African language schools.

Note that the African language group together with the group that took the test in English but are predominantly non-English home language speakers collectively comprises approximately 78% of the South African sample in PIRLS. This is closer to proportion of South African schools one would expect to come from the historically black school system. Not also that the “English: 25% or less” group of schools is substantially poorer judged by the SES index than the rest of the schools that took the PIRLS test in Afrikaans or English. Thus it may be that this distinction is salient for the purpose of improving the identification of schools into historically disadvantaged predominantly black schools and historically white, coloured and Indian schools.

Another plausible method for better understanding the distribution of reading scores amongst the "Afrikaans or English" schools is to divide this group into two halves on the basis of SES. The assumption prompting this distinction is that the bottom half probably includes mostly historically black schools, and that the top half probably includes mostly historically privileged schools. Figure 12 below shows kernel density curves of the South African reading scores by two SES halves of "Afrikaans or English" schools and by SES quintiles of African language schools.

Figure 12: Kernel density curves by language and SES

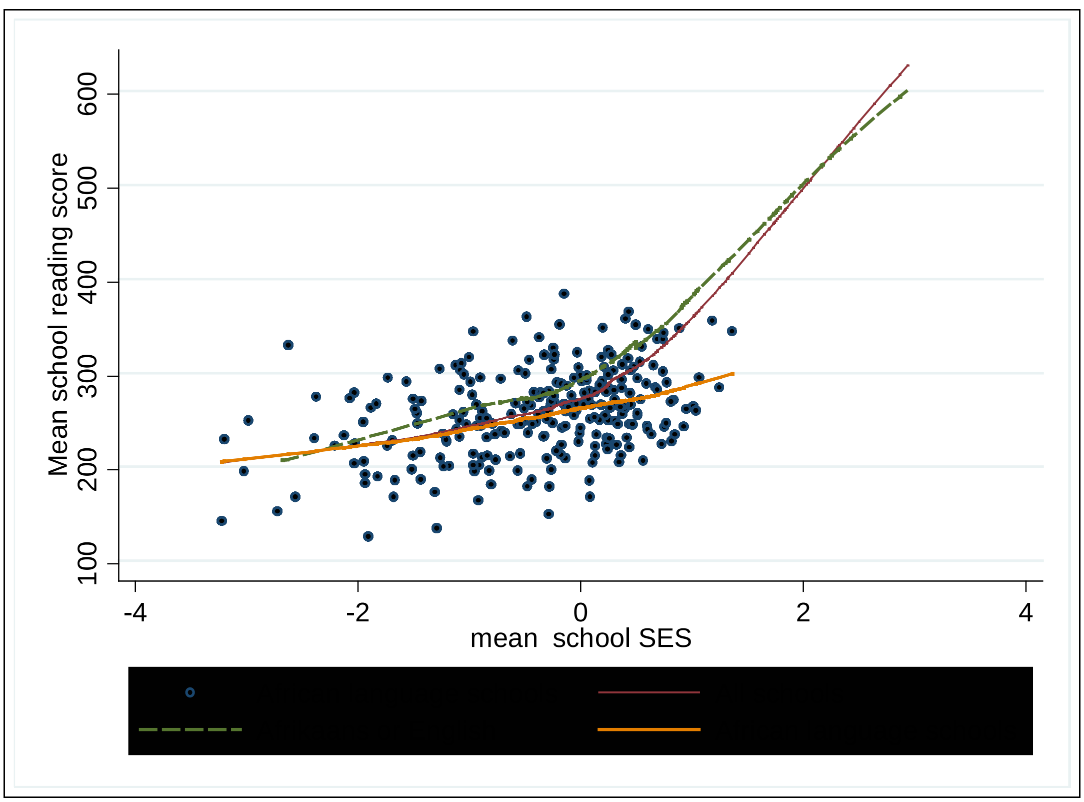


The distributions of reading scores for each African language quintile and the bottom half of "Afrikaans or English" schools are fairly similar, while the distribution for the top half of "Afrikaans or English" schools lies far to the right. This may indicate that the assumption that the bottom half of the "Afrikaans or English" schools includes mostly historically black schools and that the top half includes mostly historically privileged schools is fairly accurate. Alternatively, Figure 12 could be interpreted to show that within the historically black system improvements in SES are not associated with very substantial improvements in reading achievement, whereas in the historically white, coloured and Indian schools differences in SES are associated with widely different reading achievement.

Taken together, Figures 10, 11 and 12 contribute to the growing realisation that the distribution of educational performance in South Africa is bimodal. The classification on the basis of historically different systems seems to be appropriate for understanding this bimodality. Great inequality of educational outcomes persists despite increased equity in educational spending since political transition.

Figure 13 investigates these issues further by building on the Lowess regression analysis introduced earlier (Figure 9). Here, the original Lowess regression for South Africa (in Figure 9) is presented together with Lowess regressions by the language of the test. A scatterplot for just the "African language" schools is included for its descriptive power.

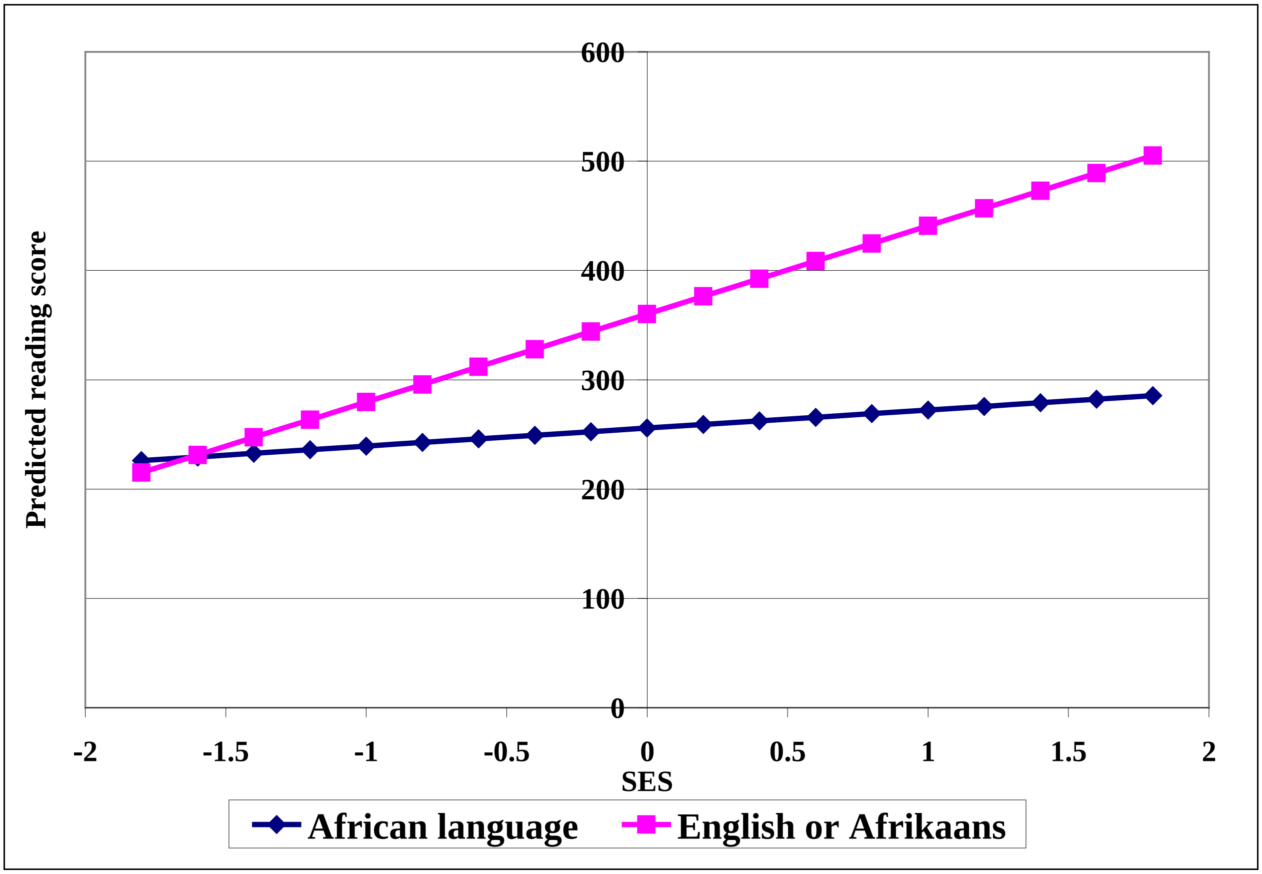
Figure 13: Lowess regressions by language of the test



The Lowess curve for the entire sample of schools is relatively flat at low levels of school SES, but then steep at high levels of SES. This could be understood as confirmation of the result from the multivariate analysis of a convex relationship between SES and reading achievement (the coefficients on both SES squared and school SES squared were positive and significant). However, it appears that "Afrikaans or English" schools account entirely for the steep section of the Lowess curve for the overall sample. In contrast, the Lowess curve for the "African language" schools remains relatively flat at all levels of school SES. The scatterplot for "African language" schools is striking because it reveals that every single such schools achieves an average reading score below the low international benchmark of 400. Moreover, there are no such schools beyond a certain threshold level of mean SES. This may relate to the phenomenon that as an emergent South African black middle class grows they are withdrawing their children from the historically black school system and placing them in historically white, coloured and Indian schools (Soudien, 2004: 107).

The difference in the effect of SES on educational outcomes between African language schools and Afrikaans/English language schools revealed by the Lowess regressions here, prompts a comparison of SES gradients using the original technique of OLS regression for these two subdivisions of the South African sample of schools. Figure 14 shows the OLS regression SES gradients for the two categories of schools.

Figure 14: SES gradients for African language schools and Afrikaans/English schools



The SES gradients confirm what the Lowess regressions indicated, that there appears to be a strong relationship between SES and reading scores amongst formerly white, coloured and Indian schools, but that amongst the historically black schools differences in home SES have little effect on reading achievement. Not too much should be made of the gradient for "Afrikaans or English" schools as there is probably a fair amount of misclassification in this group. However, there are unlikely to be any classification errors in the African language group. Therefore, it seems that for those students within the historically black school system the influence of home SES is weak.

Another observation that has bearing on this is made by Van der Berg and Louw (2006): While roughly one in ten white children of the matric-aged cohort achieved A-aggregates in the 2003 matric examination, only one in a thousand black children did, and about half of the latter group were in formerly white-only schools. Conversely, children from middle-class families rarely perform well outside of rich schools. Thus one might conjecture from Figures 13 and 14 that children in historically black schools, which constitute the bulk of the South African school system, are achieving very low educational results, while those who can afford to are escaping historically black schools and entering formerly white, coloured and Indian schools where they will receive a better quality education. Therefore, home SES appears critical in determining which school system a student enters. Then for those in the historically black system the chances of achieving high quality educational outcomes are small, regardless of home SES.

Thus the nature of educational inequality in South Africa reflects the overall nature of income inequality – it is vast and has strong racial and increasingly class dimensions. However, high levels of inequality are somewhat less troubling from a social justice point of view in the presence of considerable opportunity for social mobility. For example, according to John Rawls (1971), one of the most influential social justice theorists of the Twentieth Century, inequality is to be tolerated as long as it is to the advantage of the worst off in society. What then are the prospects for social mobility implied by the PIRLS results?

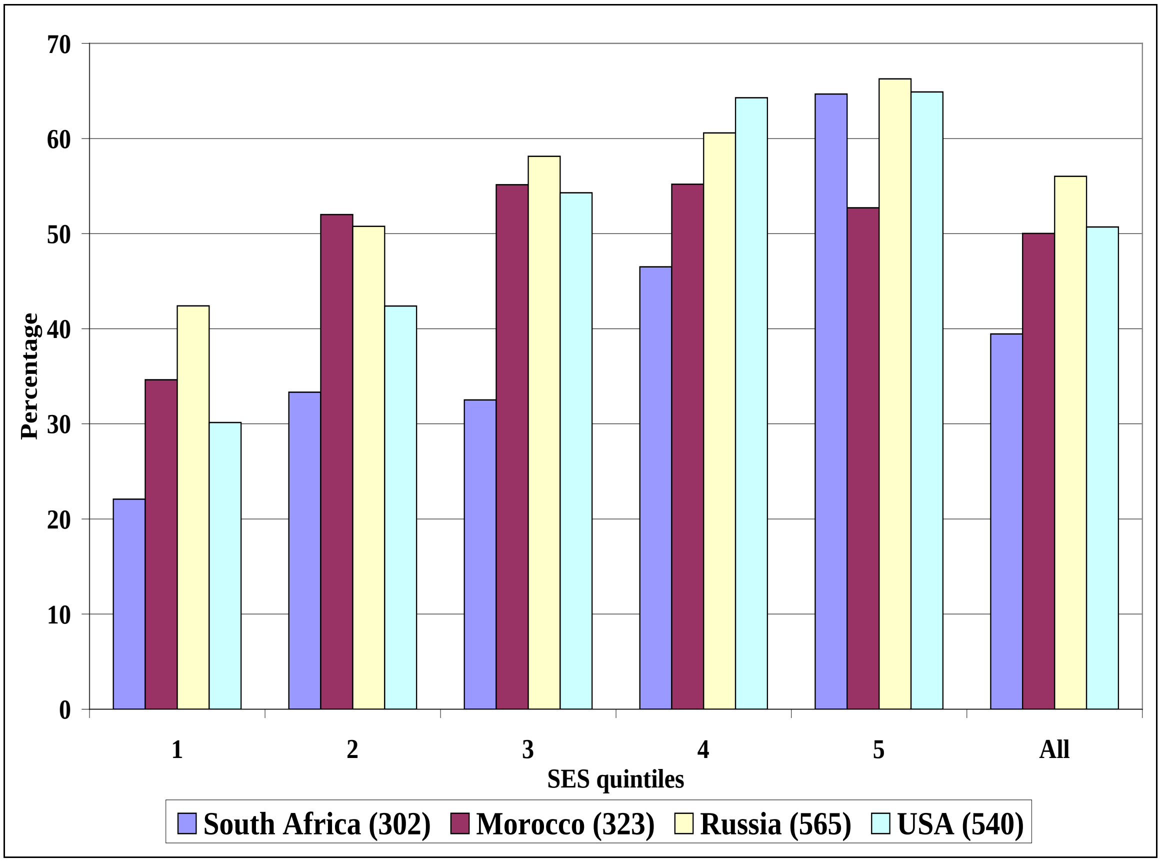
5.2 Prospects for social mobility implied by PIRLS reading scores

In this section the underlying assumption is that higher scores in PIRLS can be expected to yield returns in the form of better employability and higher earnings on the labour market in the future. This assumption relies on the logic that better educational quality in grade 5 is likely to positively influence the ultimate educational attainment of students. It also relies on the well-established finding from the earnings functions literature that both educational quality and attainment are positively correlated with employability and earnings (for example, Hanushek and Woessman, 2007). With this link between education and income in mind, this section considers the prospects for social mobility in South Africa relative to Russia, Morocco and the USA, based on PIRLS.

Figure 15 depicts the proportion of students scoring above the national average reading score in each country, by SES quintile[26]. Note that the average reading scores are 302 for South Africa, 323 for Morocco, 540 for the USA and 565 for Russia.

[26] Once again the SES index for each country was derived using Principal Components Analysis on the various "possessions" questions administered in PIRLS.

Figure 15: Proportion of students scoring above national average by SES quintile



For quintiles 1-4 South Africa has the lowest proportion of student's scoring above national average.[27] In South Africa, only 22% of students within the bottom SES quintile manage to score above the national average (302). This is very different from Morocco (a country of similarly poor reading scores and significantly lower GDP *per capita* than South Africa) where 34% of students within the bottom SES quintile score above the national average (323). It is striking that in Russia, where the national average is as high as 565, approximately 42% of students within the bottom SES quintile score above it. If reading scores in PIRLS are assumed to be a reasonable predictor of future earnings, then the results in Figure 14 suggest that the prospects for social mobility are bleak in South Africa by international comparison.

Figure 16 below depicts the proportion of students within each quintile of school mean SES that scores above the national average.[28] A similar pattern emerges to what was seen in Figure 15. For the bottom four quintiles of school mean SES, South Africa has the lowest proportion of

[27] Note that just under 40% of the entire South African sample score below the national average, indicative of the fact that the South African distribution of reading scores is skewed to the right.
[28] As before, school mean SES was calculated by taking the mean of the SES index for all students in a school.

student's scoring above national average. It is striking that in Morocco almost exactly twice the proportion of students within the bottom socio-economic quintile of schools score above the national average than is the case in South Africa. The proportion of students within the lower quintiles scoring above national average in Figure 16 are generally somewhat lower than the proportions in Figure 15, thus confirming what was established in Sections 3 and 4, namely that the combined SES of a student's school has a stronger influence on reading score than that student's own SES. This is especially so in South Africa.

Figure 16: Proportion of students scoring above national average by school mean SES quintile

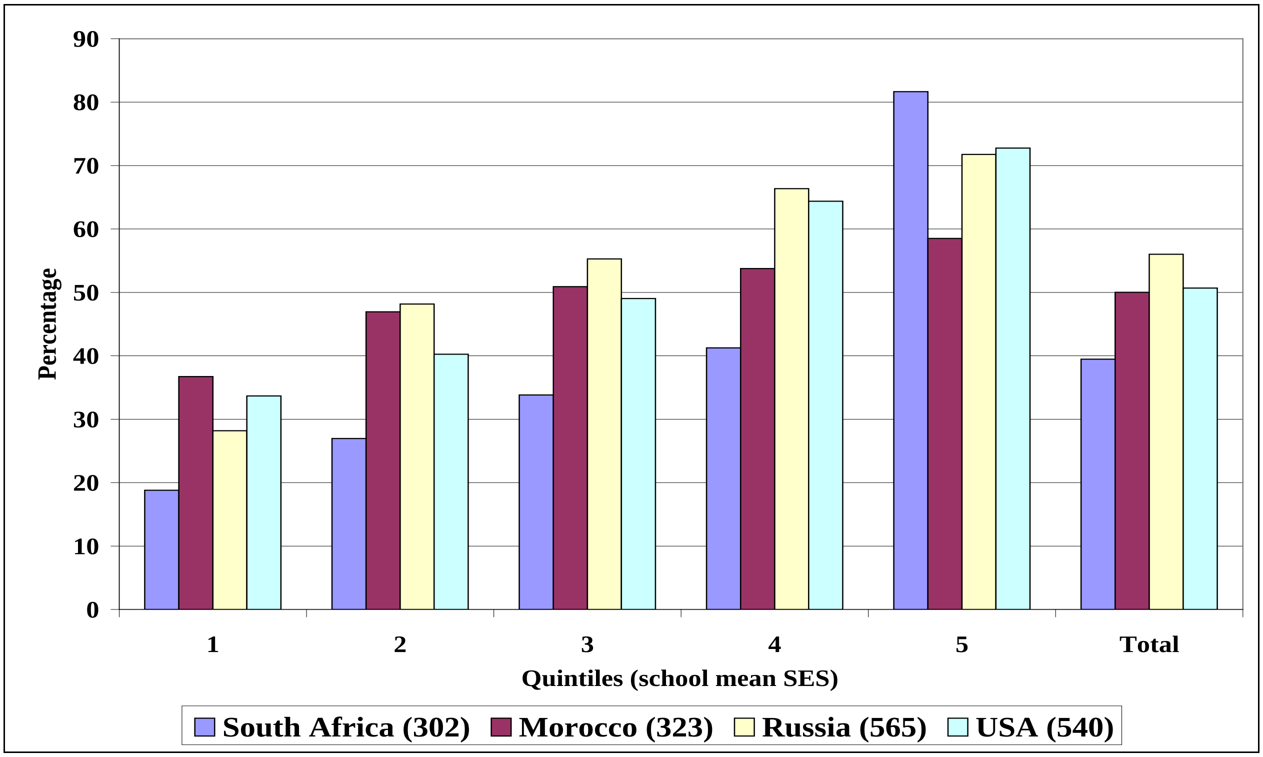
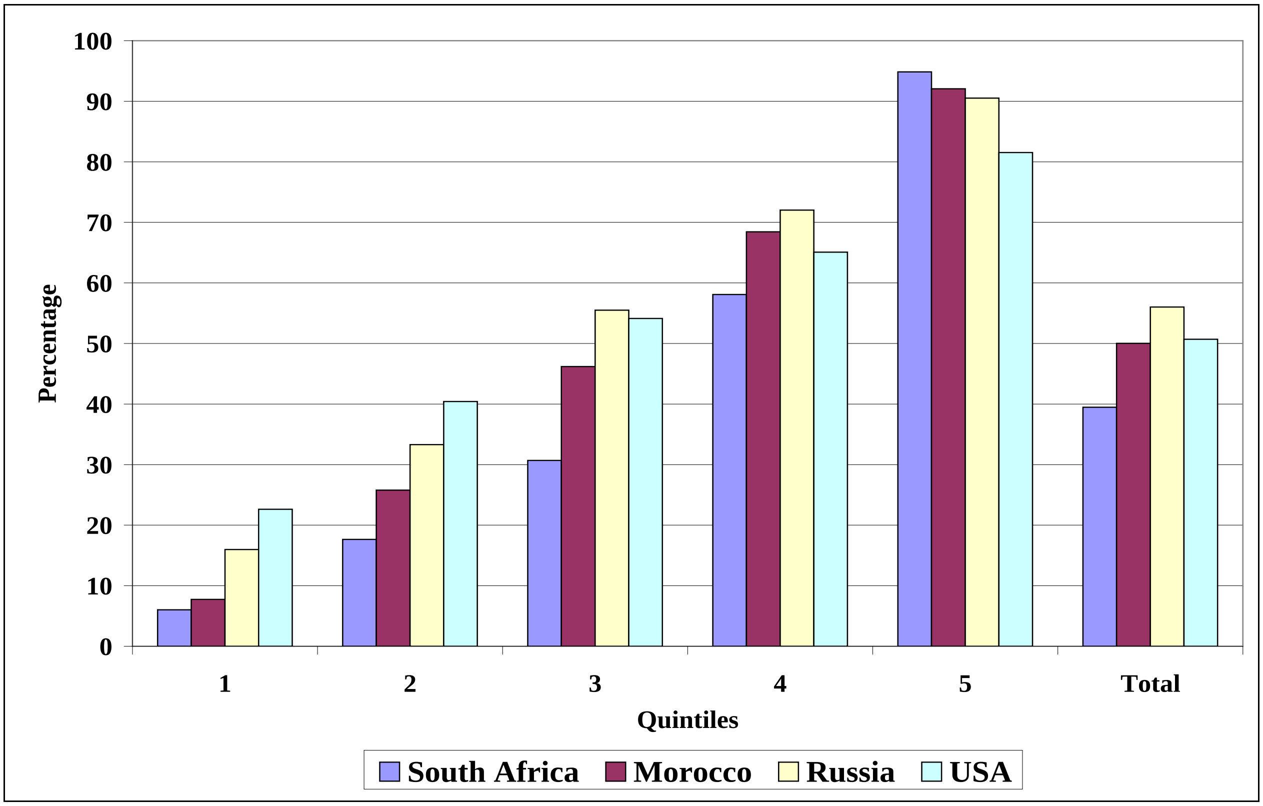


Figure 17 below presents another way of looking at the effect of schools on reading scores. Here students are divided into quintiles according to the average reading score within their school. The graph shows the proportion of students in each country scoring above the respective national averages by quintile of school mean performance. By design, therefore, one expects lower proportions of students to score above the average in the bottom quintiles. The same pattern emerges as in Figures 15 and 16. For the bottom four quintiles of school mean performance, South Africa has the lowest proportion of student's scoring above national average. Just 6% of students in the worst performing fifth of South Africa's schools manage to achieve a reading score above the national average (302), which is very low itself. This graph demonstrates that relative to students in other countries, it is harder for South African students in low quality schools to overcome the poor quality of that school and still achieve an above average reading score. Given that there is a strong correlation between school quality and school mean SES, the results in Figures 16 and 17 would suggest that the South African school system affords very limited prospects for social mobility by international standards.

Figure 17: Proportion of students scoring above national average by quintile of school mean performance



In order to gain a more nuanced perspective on the prospects for social mobility implied by the PIRLS results, it is worth searching for some benchmark reading score that can be expected to be high enough to perform on the labour market. Most estimates of the unemployment rate in South Africa, according to the broad definition, are in the region of 40%. Assuming the reading score in PIRLS is a useful predictor of future employability, and that the demand for labour will remain similar, one might suppose that the top performing 60% of South African students in PIRLS can reasonably expect to perform on the labour market in the future. The reading score at the 40th percentile of South African students is 247. This offers a very crude (and extremely low) guideline score below which students can be expected to be excluded from the future labour market, and above which students might be expected to achieve some level of success.

Figure 18 presents a similar analysis to that in Figures 15, 16 and 17. However, now the proportion of students scoring above the unemployment rate guideline score of 247 (instead of the national average reading score) is presented by quintile of student SES and by school mean SES.

<u>Figure 18: Proportion of students scoring above unemployment rate guideline (247) by student SES and by school mean SES</u>

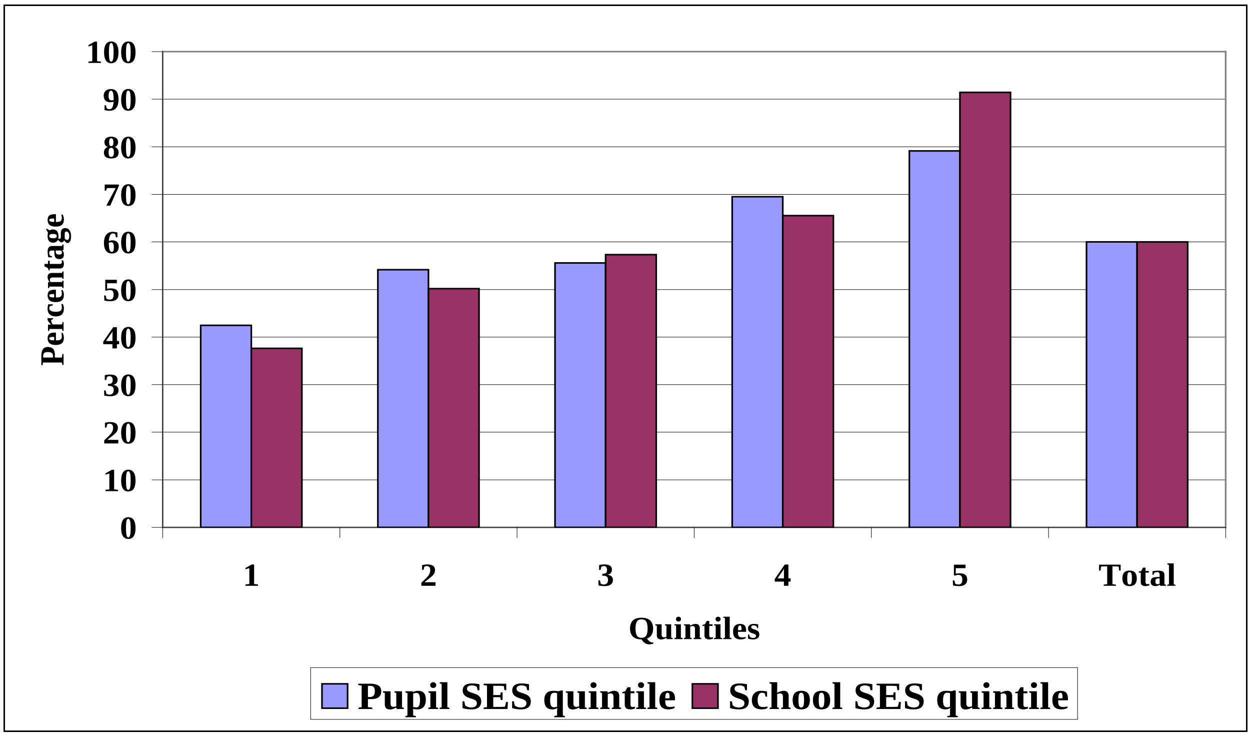


Figure 18 suggests that unemployment will be widespread amongst those groups who come from low SES backgrounds. In particular, most students in the poorest 40% of schools are likely to struggle to find employment. These results indicate that there is very limited opportunity for those with poor backgrounds to escape poverty through success on the labour market. Instead of transforming patterns of inequality, it would appear that the South African education is an institution by which existing patterns are being reproduced.

In summary, instead of schools being institutions where social mobility emanates from, it appears that in South Africa, more than elsewhere, schools are institutions that restrict social mobility.

6. Discussion

Two major findings have emerged throughout this paper. First, student background (SES) explains a great deal of the variance in reading scores amongst South African students – substantially more so than in other countries. Not only does SES explain a large amount of the variance in reading scores, but the variance is particularly great across different SES groups, indicative of the high level of educational inequality that persists in South Africa. Second, the results presented here have highlighted the importance of school-level effects, especially school mean SES. This warrants a close focus on schools as institutions that potentially amplify the effects of SES.

The finding that school mean SES is a more important determinant of reading scores than the SES of students themselves is in line with Coleman's conclusion that "the social composition of the student body is more highly related to achievement, independent of the student's own social background, than is any school factor" (quoted in Kahlenberg, 2001). It would appear from our research that the social composition of the school is in fact more important than even the student's own background, although of course the student's own SES is the major factor in determining which school they attend.

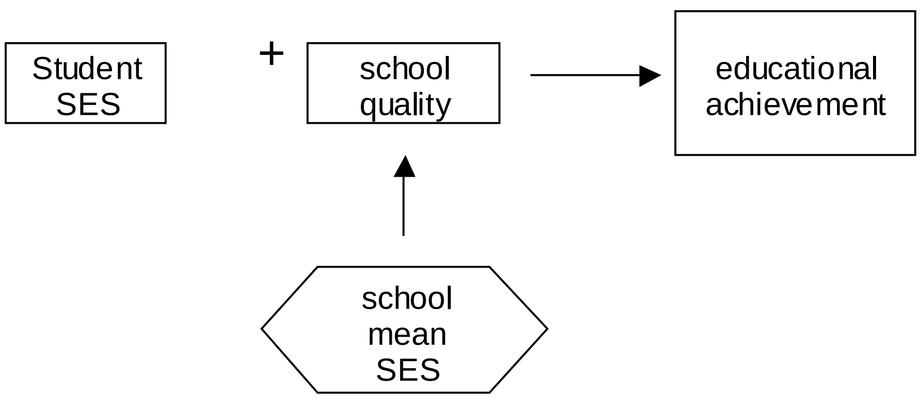
The importance of socio-economic integration was also highlighted by Willms (2004) in explaining differences in literacy scores between Canada and the USA. Willms constructed SES gradients for these two countries and found that the gradient for the USA was significantly steeper than that of Canada. The strength of the relationship was also greater for the USA. The main explanation proposed by Willms related to differences in the socio-economic intakes of US and Canadian schools. Willms (2004: 26) found that students with low SES fared better in Canada than in the USA, and that this was because of greater socio-economic integration in the schools that low SES students attend in Canada than in the USA.

Coleman asserted that not only was the SES of school peers the most important factor in determining student achievement, but it was also the very thing that was most unequally distributed across the school system in the USA. This observation is certainly truer of contemporary South Africa than it was of the USA in the sixties. What is particularly concerning is that the distribution of the SES of fellow students is becoming even more unequal. There is a movement from racial segregation in schools to more racial integration in previously white, coloured and Indian schools, but there is negligible racial integration within historically black schools, and increasing concentration of SES in the school system, rather than socioeconomic integration. Unfortunately, racial integration is less important than socio-economic integration for educational achievement, according to the Coleman Report (Kahlenberg, 2001).

It appears that class is displacing race as the critical factor in the determination of the composition of South Africa's schools, as Soudien (2004: 105) argues. In his argument, Soudien (2004: 105) uses as a point of departure the critical framework provided by Althusser, who held that ruling class ideology was transmitted through "ideological state apparatuses", such as the education system. Soudien argues that the way in which racial integration is occurring in South African schools serves to consolidate the position of the dominant class, which comprises mainly whites and, increasingly, a black middle class elite. This process occurs through the phenomenon of black flight out of historically black schools, by those who can afford it. But it also occurs more subtly through what Soudien (2004: 109) calls a "class settlement" involving the construction of a unified identity around values such as "good schooling" and maintaining quality. Although such values are in and of themselves admirable values, they are leading to the erection of barriers to entry for poor students (through higher school fees) and increasing the concentration of SES within schools.

Harley and Wedekind (2004) offer another explanation of how schooling in South Africa is serving the interests of the middle class, but not the majority of schools. They argue that Curriculum 2005, which has been implemented since 1997, was intended to be radically transformational but has had the unintended consequence of social reproduction instead. Harley and Wedekind (2004: 207) maintain that Curriculum 2005 is by its design a resource-hungry curriculum, and therefore is suited only to the middle class schools. They argue that the pupil-centred progressive ideology of Curriculum 2005 emanates from a Western individualistic culture and leads to teacher methods that are ill-suited to situations of large classes, less well educated teachers who cannot afford to neglect the expertise of textbooks, and a shortage of advanced learning support technologies. The contention that the curriculum is inappropriate for low SES schools may be a relevant perspective to keep in mind when considering how widely educational achievement varies by SES, as this paper has demonstrated.

The observation that the trend in South African schools is away from rather than toward socio-economic integration has sobering implications for one of the major questions throughout this paper – to what extent is the South African school system transforming or reproducing existing patterns of inequality? Much hope is often placed in schools as institutions of transformation. To borrow a metaphor from the then Minister of Schools Standards in the United Kingdom, David Miliband, schools should help children "climb up the down escalator", which they are on owing to their socio-economic disadvantage (quoted in New South Wales Department of Education and Training, 2005: 9). However, it would appear that in some instances, such as in South Africa, schools exert an extra downward pull keeping those at the bottom near the bottom. It would seem that the strength of this "downward pull" is linked to the high concentration of low SES in poor schools. Based on the findings in this paper, one might propose the following schematic diagram to crudely represent how student SES combined with the quality of the school they attend feeds into their educational achievement:



Note that in this model, school quality is a function of the combined SES of the students in a school. The SES of fellow students should actually be considered an important school resource, perhaps the most important, as Coleman argued. Of course this is a very crude schematic diagram and is not intended to be comprehensive. The many variables included in the multivariate analysis in Section 4 could be included. Another factor that of course feeds into educational achievement and was not available for the regression analysis is innate ability or intelligence. This is known to have a major impact on educational achievement as well as directly on labour market performance. It is striking that the full model in the regression analysis, regression [8], produced an R-squared value of 0.72 indicating that 72% of the variance in reading scores is explained by the variables in the model. This is extremely high given that intelligence is not accounted for in the model. It is worth taking a moment to theorise about how intelligence might interact with SES and educational achievement.

In a hypothetical meritocratic society, and assuming a fairly equal initial distribution of income and wealth, we could expect a process of socio-economic stratification to occur over time on the basis of intelligence or innate ability. Given that intelligence is largely genetically inherited, we would not expect a large proportion of highly intelligent children to be found amongst the poorest groups within a mature meritocracy. (Because a great deal of socio-economic stratification would already have occurred on the basis of intelligence/ability.)

In South Africa the situation is quite the opposite. Socio-economic stratification has taken place predominantly on the basis of race rather than on merit. One might therefore expect to find a sizable proportion of high-ability children emanating from poor backgrounds. However, there is no evidence from PIRLS or other similar surveys to support this expectation. An extremely small proportion of learners from the bottom quintiles score above the national average. This is probably indicative of the extent to which the family background of learners together with the combined SES of schools act as a barrier to educational achievement, and to the realization of innate ability, which may well be there. This dynamic ensures that socio-economic stratification will be reproduced instead of transformed on the basis of intelligence or innate ability.

This hypothesis may speak to the rather limited prospects for social mobility implied by the distribution of reading scores in PIRLS, as discussed in Section 5. Research by Lam (1999) and by Keswell and Poswell (2002) on the structure of labour market returns to education in South Africa is also relevant to social mobility considerations.

Lam (1999) explored the relationship between schooling inequality and income inequality in South Africa and Brazil, generally considered two of the most unequal societies in the world. Using earnings functions, he came to the unsurprising finding that schooling explains a great deal of the variance in earnings in both countries. More interestingly, Lam considered that a reduction in schooling inequality would not necessarily lead to a similar reduction in income inequality. This perhaps surprising assertion is in fact a logical implication of convexity in the returns to education. Whereas the traditional human capital model predicts diminishing marginal returns to education, Lam (1999) found the opposite to be true. There are very low returns to years of primary education and strongly increasing marginal returns at higher levels of education. This empirical finding of a convex relationship between schooling and earnings in South Africa has since become well established (e.g. Keswell and Poswell, 2002).

As Lam (1999: 23) argues, a large improvement in mean educational attainment will not necessarily affect the variance of schooling substantially. Appreciable returns on the labour market will still only be enjoyed by those at the top end of the educational distribution, leaving income inequality chiefly unaffected. Lam (1999: 23) observes that South Africa has had substantially less educational inequality than Brazil, but that this has not translated into a more equal income distribution. This is chiefly because of the convex structure of returns to education.

Similarly, Keswell and Poswell (2002: 20-21) consider some of the implications of convex returns to education for inequality. In the South African context, where there is a significant cost attached to acquiring education (especially at the higher levels), disparities in household wealth (and SES) determine who is able to reach the level of education that is high enough to be rewarded on the labour market. One implication of this scenario is that talented individuals from poor backgrounds might make a fully rational decision to drop out of school fairly early on, simply because the opportunity cost of attending primary and most of secondary school is greater than the returns such levels of schooling are likely to secure. With a structure of convex returns to education, the rich have a greater incentive to acquire education than do the poor, thus reinforcing inequality. Given this convexity, Keswell and Poswell (2002: 20-21) suggest that educational reforms and small-scale interventions are likely to have a negligible impact on income inequality.

The research by Lam (1999) and by Keswell and Poswell (2002) is restricted to analysing the returns to educational attainment, and thus does not account for the quality of education received. A number of studies investigating differential earnings by race in South Africa have pointed to the problem of poor education quality in predominantly black schools, e.g. Van der Berg (2001: 179) and Burger and Jafta (2006). These two studies find that a racial wage gap persists even after educational attainment and work experience have been accounted for. The authors speculate that differentials in education quality lie behind this result.[29] This paper confirms that

[29] One might also speculate that unobserved factors associated with SES, such as productivity enhancing characteristics learnt at home (e.g. Business principles learnt from a parent that owns a business) and non-productive

there are wide differences in quality across the South African school system. The disparities in the quality of education being received by students with different SES only worsen the already bleak outlook for social mobility implied by the findings of convex returns to education attainment. Given that the quality of schooling is an important determinant of attainment, and that high levels of attainment are required before labour market rewards become significant, the need arises for improving educational quality to be the focus of attention in schools where low SES is concentrated. This brings us to some implications for policy.

Before school interventions are considered, it should be noted that early childhood support to low SES families should be viewed as an important domain of education policy, as there is strong evidence that low SES students enter school with a backlog in cognitive development (e.g. Lee and Burkam, 2002).

One group of school-level policy options include targeting SES and its effects directly. Schemes to assist in transporting children to school in cases where long distances increase the incidence of absenteeism and shorten the effective school day, as well as school-feeding programmes can help overcome the disadvantages associated with low SES. Progressive spending on education is certainly one policy option. There has already been substantial progress towards equity in educational spending since the transition, and more recently the "no fee" schools policy represents a significant move to remove barriers to schooling. However, there is a growing realisation that increased equity in education spending is not translating into equity in outcomes (e.g. Van der Berg, 2002; Fiske and Ladd, 2004). That said, there is reason to believe that well targeted spending on school resources can be effective. For example, Behrman *et al* (1998) analyse the effects of school policies on intergenerational mobility in Latin America, and find that increased spending on resources at primary school level can have a positive impact on mobility especially when it is aimed at improving school quality, although increased spending on tertiary education can have a regressive effect. The lesson is that spending *per se* will not guarantee returns, but that carefully targeted spending on items and structures, for which there are empirical reasons to expect them to lead to school improvement in the case of South Africa, may significantly benefit low SES and under-performing schools.

---

characteristics such as access to social networks account for some of the unexplained component of the racial wage gap.

The fact that spending on resources does not seem to guarantee improved outcomes leads to a focus on policies that address the internal efficiency of schools. Efficiency problems within many of South Africa's poor schools include student and teacher absenteeism, the ineffective use or the non-use of resources such as textbooks, low work ethic, etc. Interventions that aim to sharpen the monitoring and accountability structures facing teachers, through *inter alia* fostering greater parent involvement and awareness of what goes on in schools, a factor which is very often lacking in schools with predominantly low SES student bodies, may be necessary. Such policies target the types of problems that tend to be associated with schools and communities with a high concentration of low SES.

Another policy option that is only logical given the findings of this paper is to promote socio-economic integration within South Africa's schools. This was precisely the major motivation behind the decision at the time of political transition to allow schools to retain the right to charge fees. It was felt that if public schools were not allowed to charge fees this would prompt a mass exodus of high SES (mainly white) students from the public school system to private schools, with detrimental consequences for the quality of public schools. A key policy recommendation of the Coleman Report was that incorporating low SES students into schools with a critical mass of higher SES peers would significantly reduce inequalities of educational opportunity in the USA. Unfortunately, the sheer scale of poverty in South Africa means that there just is not the capacity for wealthier schools to absorb the necessary numbers of poor students and still retain a critical mass of high SES students. This is not to say that more innovative and probably incremental interventions targeting the social composition of schools in South Africa are not worth exploring. Certainly the phenomenon of middle-class flight from the historically black school system, with a consequent added negative effect on quality, needs to addressed as the majority of the country's students remain trapped in this failing part of the school system.

Lastly, the concerns surrounding the unintended consequences of the curriculum warrant attention. Fortunately this is beginning to be recognised by policy-makers as a legitimate concern, whereas for too long an almost religious faith was invested in the transformatory promise of Curriculum 2005. Ultimately, however, the search for policy solutions is beyond the scope of this paper.

Conclusion

The analysis of South Africa's performance in PIRLS 2006 yields three very concerning results. First, the overall performance measured by the national average reading score is extremely low by international standards. Second, SES is a very important determinant of reading achievement, both in terms of how widely reading achievement varies by SES and how much of the overall variance in reading scores is explained by SES. Third, students in poorly performing and low SES schools find it especially hard to overcome this disadvantage, a state of affairs that does not bode well for social mobility.

The question as to whether education systems in general can transform inequalities or merely end up reproducing them remains open. However, the above findings indicate that given the combined set of circumstances currently prevailing in South Africa, the school system is unlikely to make a significant contribution to social mobility. In this respect, the school system may be acting as a constraint on South Africa's economic development. It is not inevitable that this unhappy situation will persist indefinitely. However, meaningful change will require well-targeted innovative solutions and a strong political commitment sustained over a long period of time.

Bibliography

Anderson, K. G. 2000. 'Family Structure, Parental Investment, and Educational Outcomes among Black South Africans,' *Population Studies Center, University of Michigan*. Research Report No. 00-461.

Anderson, K. G., Case, A. and Lam, D. 2001. 'Causes and Consequences of Schooling Outcomes in South Africa: Evidence from Survey Data.' *Social Dynamics.* 27(1): 1-23.

Barro, R. J. and Lee, J. 1997. 'Schooling Quality in a Cross Section of Countries,' NBER Working Paper, No. 6198.

Behrman, J. R., Birdsall, N. and Székely, M. 1998. 'Intergenerational Schooling Mobility and Macro Conditions and Schooling Policies in Latin America.' Inter-American Development Bank, Working Paper No. 386.

Bullock, A. and Stallybrass, O. (eds.) 1982. *The Fontana Dictionary of Modern Thought*. London: Fontana/Collins.

Burger, R. and Jafta, R. 2006. 'Returns to race: Labour market discrimination in post-apartheid South Africa.' Stellenbosch Economic Working Paper 04/06.

Filmer, D., and Pritchett, L. 1999. 'The Effect of Household Wealth on Educational Attainment: Evidence from 35 Countries.' *Population and Development Review* 25(1): 85-120.

Filmer, D., and Pritchett, L. 2001. 'Estimating Wealth Effects without Expenditure Data – or Tears: An Application to Educational Enrollments in States of India.' World Bank Working Paper.

Fiske, E.B. and Ladd, H.F. 2004. *Elusive Equity: Education Reform in Post-Apartheid South Africa.* Washington D.C.: Brookings Institution Press.

Hanushek, E. and Woessmann, L. 2007. 'The role of school improvement in economic development.' NBER Working Paper, No. 12832.

Harley, K. and Wedekind, V. 2004. 'Political change, curriculum change and social formation, 1990 to 2002. pp. 195-220. In, Chisholm, L. (ed.) 2004. *Changing Class: Education and Social Change in Post-Apartheid South Africa*. Cape Town: HSRC Press.

Kahlenberg, R. D. 2001. 'Learning from James Coleman.' *Public Interest*. Summer, 2001.

Keswell, M. and Poswell, L. 2002. 'How Important is Education for Getting Ahead in South Africa?' CSSR Working Paper No. 22. December 2002.

Kolenikov, S. and Angeles, G. 2004. 'The Use of Discrete Data in PCA: Theory, Simulations, and Applications to Socioeconomic Indices.' Carolina Population Center.

Lam, D. 1999. 'Generating Extreme Inequality: Schooling, Earnings, and Intergenerational Transmission of Human Capital in South Africa and Brazil.' *Population Studies Center, University of Michigan*. Research Report No. 99-439.

Lee, V. E. and Burkham, D. T. 2002. *Inequality at the Starting Gate: Social Background Differences in Achievement as Children Begin School*. Washington, D.C.: Economic Policy Institute.

Mankiw, N. G., Romer, D., and Weil, D. 1992. 'A Contribution to the Empirics of Economic Growth.' *Quarterly Journal of Economics*. Vol.107(May). 407-37.

Meier, G.M. 1995. *Leading Issues in Economic Development.* 6th Edition. Oxford: Oxford University Press.

Mullis, V. S., Martin, M. O., Kennedy, A. M. and Foy, P. 2007. *PIRLS 2006 International Report*. Boston: TIMSS & PIRLS International Study Center.

New South Wales Department of Education and Training. 2005. *Review of the Recent Literature on Socio-economic Status and Learning*. Erebus International.

Raudenbush, S. W. and Bryk, A. S. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. London: Sage Publications.

Rawls, J. 1999(1971). *A Theory of Justice.* Oxford: Oxford University Press.

Ross, K. N. Zuze, L. 2004. 'Traditional and alternative views of school system performance,' *Educational Quality*. IIEP Newsletter, October-December, 8-9.

Sen, A. 1999. *Development as Freedom.* Oxford: Oxford University Press.

Soudien, C. 2004. ' "Constituting the class": an analysis of the process of "integration" in South African schools.' Pp. 89-114. In, Chisholm, L. (ed.) 2004. *Changing Class: Education and Social Change in Post-Apartheid South Africa*. Cape Town: HSRC Press.

Van der Berg, S. 2001. 'Social Policy to Address Poverty,' In, Bhorat, H., Leibbrandt, M., Maziya, M., Van der Berg, S. and Woolard, I. 2001. *Fighting Poverty: Labour markets and inequality in South Africa.* UCT Press: Cape Town.

Van der Berg, S. 2002. *Poverty, fiscal incidence and social outcomes.* Paper commissioned by gtz (German Agency for Technical Cooperation) on behalf of the Policy Coordination and Advisory Unit in the presidency for the ten year review of government policy. Stellenbosch: November 2002.

Van der Berg, S. and Louw, M. 2006. *Unravelling the mystery: Understanding South African schooling outcomes in regional context.* Paper to the conference of the Centre for the Study of African Economies. Oxford: CSAE, Oxford University. 21st March 2006.

Vyas, S. and Kumaranayake, L. 2006. 'Constructing socio-economic indices: how to use principal components analysis.' Oxford University Press.

Willms, J. D. 1997. 'Literacy skills and social class,' *Options Politique*. July/August. 22-26.

Willms, J. D. 2004. 'Reading Achievement in Canada and the United States: Findings from the OECD Programme of International Student Assessment.' Report to the *Learning Policy Directorate, Strategic Policy and Planning, Human Resources and Skills Development, Canada*.

Appendix A: Mean reading achievement in each country

Country	Reading overall						Reading purposes		Comprehension processes	
	Mean (All)	Std error of mean	Std Dev. (All)	Mean (Boys)	Mean (Girls)	Mean Diff. (Boys – Girls)	Mean: Literary	Mean: Informa-tional	Mean: Retrieval & Straight-forward inferen-cing	Mean: Inter-preting, inte- grating & evalua-ting
Austria	538	2.2	64	533	543	10	537	536	544	530
Belgium (Flemish)	547	2.0	56	544	550	6	544	547	545	547
Belgium (French)	500	2.6	69	497	502	5	499	498	501	497
Bulgaria	547	4.4	83	537	558	21	542	550	538	553
Canada, Alberta	560	2.4	68	556	564	8	561	556	553	564
Canada, British Columbia	558	2.6	69	554	562	9	559	554	551	562
Canada, Nova Scotia	542	2.2	76	531	553	21	543	539	533	548
Canada, Ontario	555	2.7	71	549	562	13	555	552	543	563
Canada, Quebec	533	2.8	63	527	539	13	529	533	533	531
Chinese Taipei	535	2.0	64	529	542	13	530	538	541	530
Denmark	546	2.3	70	539	553	14	547	542	551	542
England	539	2.6	87	530	549	19	539	537	533	543
France	522	2.1	67	516	527	11	516	526	523	518
Georgia	471	3.1	75	463	480	17	476	465	478	461
Germany	548	2.2	67	544	551	7	549	544	555	540
Hong Kong	564	2.4	59	559	569	10	557	568	558	566
Hungary	551	3.0	70	548	554	5	557	541	544	554
Iceland	511	1.3	68	501	520	19	514	505	516	503
Indonesia	405	4.1	79	395	415	20	397	418	409	404
Iran	421	3.1	95	414	429	14	426	420	428	418
Israel	512	3.3	99	506	520	15	516	507	507	516
Italy	551	2.9	68	548	555	7	551	549	544	556
Kuwait	330	4.2	111	297	364	67	340	327	337	n/a**
Latvia	541	2.3	63	530	553	23	539	540	534	545
Lithuania	537	1.6	57	528	546	18	542	530	531	540
Luxembourg	557	1.1	66	556	559	3	555	557	565	548
Macedonia	442	4.1	101	432	453	21	439	450	446	439
Moldova	500	3.0	69	493	507	14	492	508	486	515
Morocco	323	5.9	109	314	332	18	317	335	336	n/a**
Netherlands	547	1.5	53	543	551	7	545	548	551	542
New Zealand	532	2.0	87	520	544	24	527	534	524	538
Norway	498	2.6	67	489	508	19	501	494	502	495
Poland	519	2.4	75	511	528	17	523	515	516	522
Qatar	353	1.1	96	335	372	37	358	356	361	n/a**
Romania	489	5.0	91	483	497	14	493	487	489	490
Russian Federation	565	3.4	69	557	572	15	561	564	562	563
Scotland	527	2.8	80	516	538	22	527	527	525	528
Singapore	558	2.9	77	550	567	17	552	563	560	556
Slovak Republic	531	2.8	74	525	537	11	533	527	529	531
Slovenia	522	2.1	71	512	532	19	519	523	519	523
South Africa	302	5.6	136	283	319	36	299	316	307	n/a**
Spain	513	2.5	71	511	515	4	516	508	508	515
Sweden	549	2.3	64	541	559	18	546	549	550	546
Trinidad & Tobago	436	4.9	103	420	451	31	434	440	438	437
United States	540	3.5	74	535	545	10	541	537	532	546
International average*	500			492	509	17	500	501	501	519
High-income average*	515			507	523	17	515	514	516	529
Upper middle-income average	509			500	518	18	510	508	505	537
Lower middle-income average	427			419	436	17	425	433	431	448

Note: the international average, high-income average, upper middle-income average and the lower middle-income average are simply the (unweighted) mean of the (weighted) mean reading score of the countries.
* The five Canadian provinces are excluded and the two Belgium education systems are treated as two separate countries when the international and high-income countries' average are calculated.
** The mean score in interpreting process is not available in Kuwait, Morocco, Qatar and South Africa.

Appendix B: Kernel Density Curves

A Kernel Density estimate is a non-parametric estimator, i.e. there is no fixed functional form – every data point feeds into the estimate. Kernel Density curves represent an improvement on histograms, which are the simplest type of non-parametric density estimates.

In constructing a histogram, the data is divided into equal intervals called bins. Each time a data point falls within a particular interval a box is placed on top of that bin, increasing its height. There are three main disadvantages to histograms. They are not smooth. They depend crucially on the end points of the bins. They depend on the width of the bins. Kernel Density curves provide a solution to these disadvantages.

Kernel estimators centre a kernel function at each data point, thus overcoming the dependence on the bin endpoints. In the analysis in this paper, the commonly used Gaussian kernel function was applied. Kernel Density curves graph the results of all the kernel estimates, smoothing over the contribution of each data point relative to its local neighbourhood. The extent of the smoothing depends on the bandwidth chosen. Undersmoothing will occur when the chosen bandwidth is too narrow. Conversely, oversmoothing will occur when the chosen bandwidth is too wide, resulting in a loss of important trends in the data. The methods that can be used to determine the optimal bandwidth are not discussed here. In our analysis, we followed the procedure used in *stata* to calculate the optimal bandwidth, i.e. the width that minimizes the mean integrated squared error.

Appendix C: A note on the use of parent's education in deriving an index for SES

The educational attainment of parents is generally considered to be one of the best proxies for SES. Although PIRLS is useful in that it collected information on parent's education, including this variable in the SES index presents some methodological challenges.

The educational attainment of parent's education as collected in PIRLS is an ordinal categorical variable. In the case of South Africa there were six categories:

1) 0-9 years schooling
2) complete grade 9
3) complete matric
4) post secondary training (vocational training)
5) first degree (diploma)
6) Honours/Masters/PhD degree

In Principal Components Analysis an ordinal categorical variable such as this will be interpreted as if it were a continuous variable. One solution that has been suggested is to recode such variables into a number of binary variables entering them as separate dummies into the Principal Components Analysis (eg. Filmer and Pritchett, 2001 and Vyas and Kumaranayake, 2006). However, an important objection to this procedure is made by Kolenikov and Angeles (2004). They point out that such dummies will necessarily be negatively correlated with each other. This will "confuse" the Principal Components Analysis as to whether most of the common variation amongst the variables is caused by the correlation amongst the dummies due to the coding technique or by the unobserved factor, such as SES, which is what one aims to isolate with Principal Components Analysis (Kolenikov and Angeles, 2004: 5).

Kolenikov and Angeles (2004) then test various methods of including ordinal categorical variables in Principal Components Analysis by simulating data and testing how closely each method corresponds to the "true" values. They come out in opposition of using dummy variables and suggest that the results will be less biased if such variables are entered simply in theirordinal form. They suggest that the coding of the categories should be evenly spaced so as to avoid unnecessary bias. Another possible method is to transform the parent education variable into a pseudo continuous variable by estimating the midpoint of the years of education corresponding to each category. In the case of South Africa the following midpoints were estimated as years of education:

1) 0-9 years schooling	4 years
2) complete grade 9	9 years
3) complete matric	12 years
4) post secondary training (vocational training)	14 years
5) first degree (diploma)	15 years
6) Honours/Masters/PhD degree	16 years

There are still further challenges related to using parent's education in the SES index. The categories are different across countries where education systems are different, making comparison difficult, and in the USA the question was not even administered. Another problem is that in South Africa there was a very low response rate to this question (60%).

Despite these problems SES gradients for South Africa, Russia and Morocco were estimated using parent education for interest sake. The issue of missing data was dealt with by imputing the mode educational attainment of parents of other students in the school that a student is in. This was considered better than cutting the sample size, especially as non-response was significantly negatively correlated with reading scores. Figures 19, 20 and 21 compare the original SES gradients using only the "possessions variables" with SES gradients using the SES index that includes parent education entered as education years (the midpoint method).

Figure 19: SES gradients for South Africa

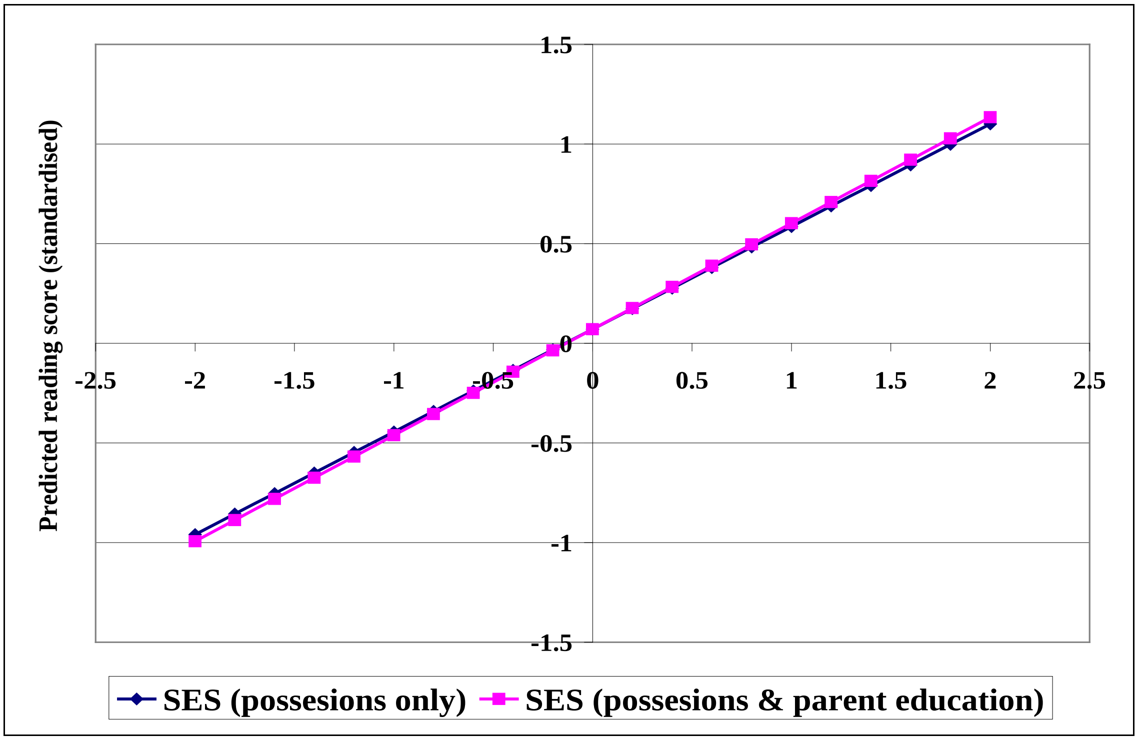


Figure 20: SES gradients for Russia

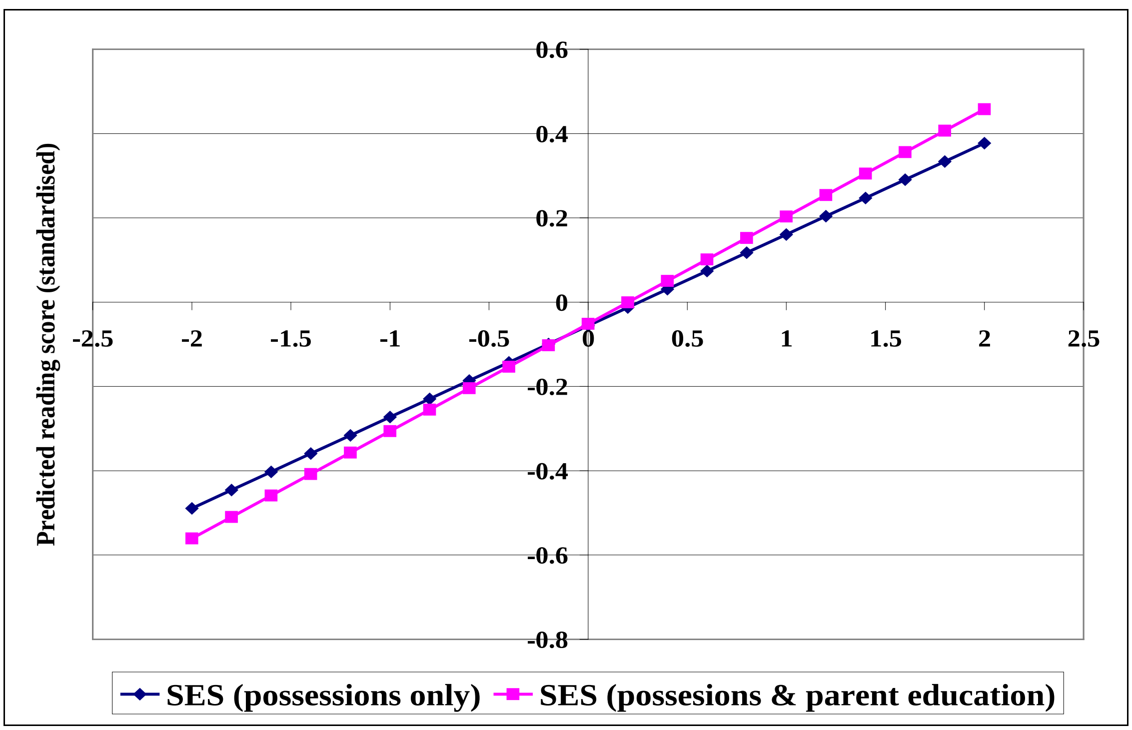
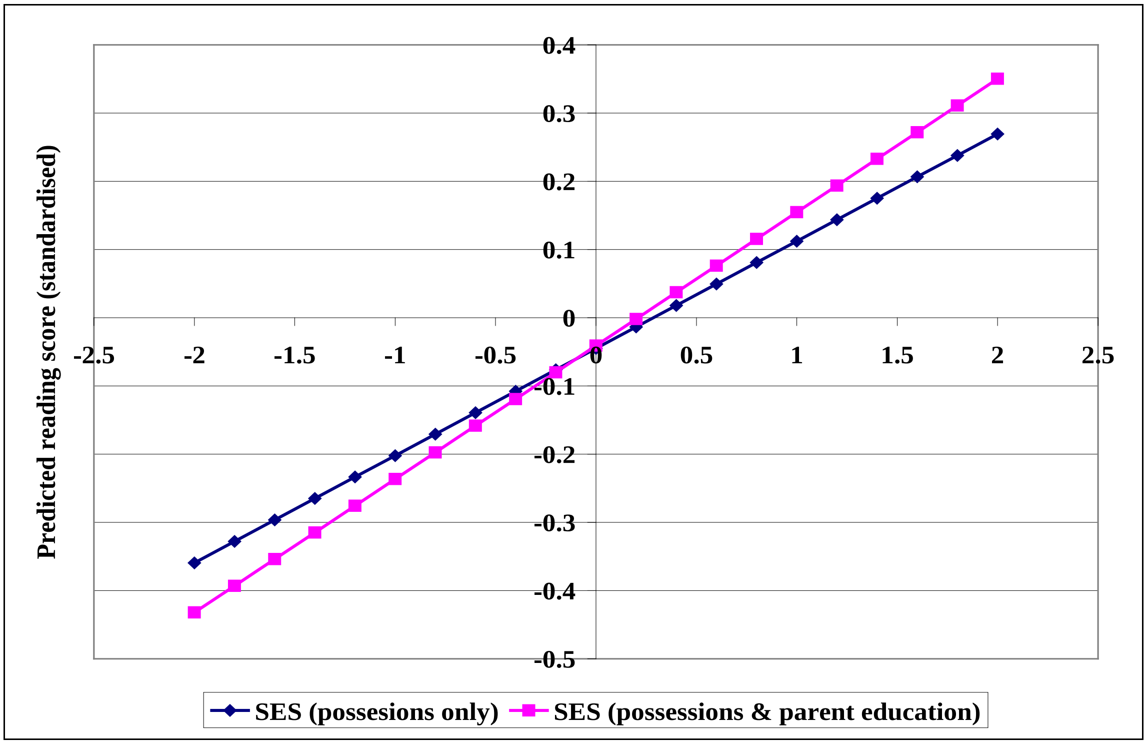


Figure 21: SES gradients for Morocco



The figures illustrate that the gradients are not substantively different when parent education is included in the Principal Components Analysis used to generate an SES index – especially in the case of South Africa. Although in the cases of Russia and Morocco, a little more of the variation in reading scores is explained by the SES index with parent education, as indicated by the R-squared statistics. For Russia, the R-squared goes up from 0.0444 for possessions only to 0.0618 when parent education is included. For Morocco, the R-squared similarly increases from 0.0232 to 0.0361.

Thus including parent's education in the index for SES does not add much to what the SES gradients constructed with only the "possessions" variables tell us. Considering all the difficulties involved in the procedure, it was decided to use the simpler SES index throughout the analysis in this paper.

Appendix D: Comparing PIRLS with SACMEQ and TIMSS

Table 5 below shows the information on the three surveys that South Africa took part in. The pupil SES was derived in all surveys using Principal Components Analysis on the "possessions" variables. The following should be noted:

- The number of items and the type of items included for the SES are slightly different in each survey.
- It is assumed that the student does not possess an item if his answer is 'don't know' or 'unspecified'.
- SACMEQ II and PIRLS2006 are the only two surveys that test on students' reading ability.
- The pupil SES and achievement scores in each subject are converted into a standardised variable with mean and standard deviation equaling 0 and 1 respectively.
-

Table 5: General information on the South African TIMSS, SACMEQ II and PIRLS data

	TIMSS1995	TIMSS1999	TIMSS2003	SACMEQ II	PIRLS2006
GENERAL INFORMATION					
Year of survey	1995	1998	2002	2000	2005
Grade of pupils	All: Gr 7 – 8	Gr 8	Gr 8	Gr 6	Gr 5
Expected birth year of pupils	1981 / 1982	1984	1988	1988	1994
Number of pupils	9 792[30]	8 146	8 952	3 163	14 657
Number of schools	234	194	255	169	397
OWNERSHIP OF ITEMS/ASSETS					
Possession items for SES	***Total: 16*** • Calculator • Computer • Study desk/table • Dictionary • Electricity • Tap water • Warm water • Radio • TV • VCR • Tape recorder • CD player • Own room • Bicycle • Flush toilet • Motor car	***Total: 14*** • Calculator • Computer • Study desk/table • Dictionary • Electricity • Tap water • TV • VCR • CD player • Radio • Own room • Flush toilet • Motor car • Bicycle	***Total: 16*** • Calculator • Computer • Study desk/table • Dictionary • Electricity • Tap water • TV • VCR • CD player • Radio • Own room • Flush toilet • Motor car • Bicycle • Telephone • Fridge	***Total: 14*** • Newspaper • Magazine • Radio • TV • VCR • Cassette player • Telephone • Refrigerator • Car • Motorcycle • Bicycle • Piped water • Electricity • Table to write on	***Total: 10*** • Computer • Study desk/table • Own books • Newspaper • Own room • Own cellphone • Calculator • Dictionary • Electricity • Tap water
No. of students specifying answers on ALL items	8 055 (82.26%)	7 012 (86.08%)	7 066 (78.93%)	3 163 (100.00%)	11 023 (75.21%)
MEAN SCORE					
Reading/Language	n/a	n/a	n/a	492	302
Maths	276	275	264	486	n/a
Science	260	243	244	n/a	n/a

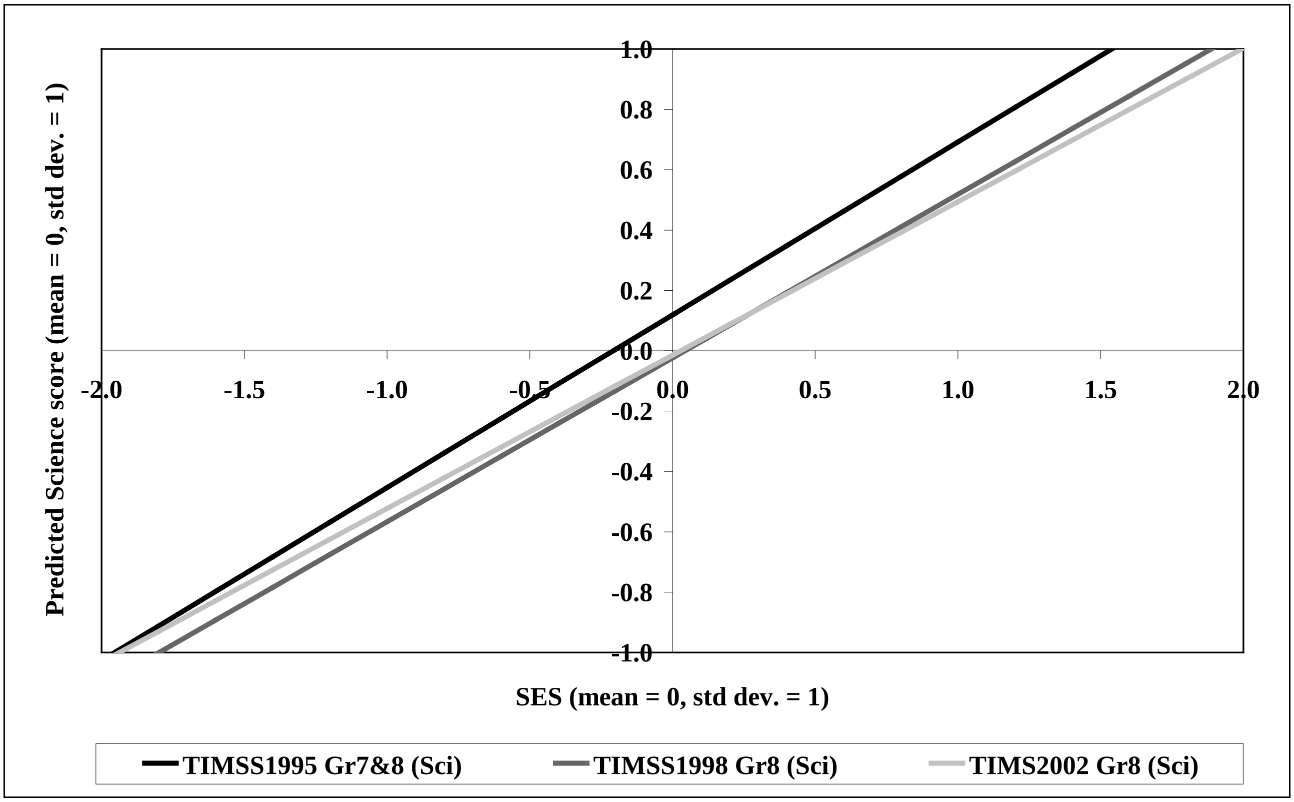
[30] The number of Grade 7 and Grade 8 students are 5 301 and 4 491 respectively.

Table 6 below presents the results on the regressions using SES and SES-squared as the explanatory variables.

Table 6: Results of the regressions

* Significant at 5%		Pupil score = f(SES)			Pupil score = f(SES, SES-squared)			
		Constant	**SES**	**R-squared**	**Constant**	**SES**	**SES-squared**	**R-squared**
READING								
RSA	**SACMEQ II**	0.0637	0.5232*	0.2609	-0.1773*	0.5341*	0.2332*	0.3072
	PIRLS 2006	0.0695	0.5169*	0.2223	-0.1567*	0.5597*	0.2166*	0.2647
MATHS								
RSA	**SACMEQ II**	0.0630	0.4428*	0.1976	-0.2021*	0.4550*	0.2550*	0.2561
	TIMSS 1995	0.1110*	0.5502*	0.2572	-0.2047*	0.5670*	0.2858*	0.3624
	TIMSS 1998	-0.0121	0.5283*	0.2794	-0.2490*	0.5913*	0.2315*	0.3545
	TIMSS 2003	-0.0067	0.4948*	0.2356	-0.3387*	0.6058*	0.3161*	0.3627
SCIENCE								
RSA	**TIMSS 1995**	0.1190*	0.5726*	0.2775	-0.2088*	0.5899*	0.2968*	0.3906
	TIMSS 1998	-0.0241*	0.5426*	0.2931	-0.2893*	0.6133*	0.2593*	0.3868
	TIMSS 2003	-0.0145	0.5085*	0.2486	-0.3368*	0.6163*	0.3070*	0.3683

Appendix E: Science score = f(SES) in South Africa



Appendix F: Deriving the intra-class correlation coefficient (Rho)

The technique of Hierarchical Linear Modelling (HLM) is appropriate for analysing data that has a nested or clustered structure. In the present situation, students are nested in classes in schools, and we may expect there to be factors common to each classroom that influence student reading achievement. Thus, HLMs can be used to “pose hypotheses about relationships occurring at each level and across levels and also assess the amount of variation at each level” (Raudenbush and Bryk 2002: 5). In the present analysis, students form level-1 and schools form level-2 of the HLM. The dependent variable is reading score and there are no explanatory variables in either level of the model, making this a fully unconditional HLM. Assuming there are j schools in the sample, the level-1 and level-2 models are given by the following:

Level-1: Reading score = $\beta_{0j} + r_{ij}$

Level-2: $\beta_{0j} = \gamma_{00} + \mu_{0j}$

Where r_{ij} is the level-1 error term, which is assumed to be normally distributed with a mean of zero and a constant variance, $V(r_{ij})$; the intercept β_{0j} is the mean reading score of the jth school and is also the dependent variable in the level-2 model; γ_{00} is the overall mean reading score for the entire sample; and μ_{0j} is the level-2 error term, with an assumed mean of zero and variance of $V(\mu_{0j})$. Note that the r_{ij} comprises deviations from the school mean reading score by each individual student, and that μ_{0j} comprises deviations of each school mean reading score from the overall mean reading score for the entire sample.

Substituting the level-2 model into the level-1 model yields:

Reading score = $\gamma_{00} + \mu_{0j} + r_{ij}$

Given that γ_{00} is the overall mean reading score, the variance of reading scores is given by:

Var(reading score) = Var($\mu_{0j} + r_{ij}$) = $V(\mu_{0j})$ + $V(r_{ij})$

As Raudenbush and Bryk (2002: 24) observe, $V(\mu_{0j})$ captures the between-school variability, while $V(r_{ij})$ captures the within-school variability. Now the intra-class correlation coefficient (rho) can be generated using the formula presented in the main text of this paper:

$$\rho = V(\mu_{0j})/[V(\mu_{0j}) + V(r_{ij})]$$

As explained in the main text, this rho value (an index between zero and one) captures the proportion of the overall variance in reading scores explained by between-school differences as opposed to within-school differences amongst individuals.

It should be noted that in our calculations using the PIRLS data, the weight variables at level 1 and level 2 are pupil weight and school weight respectively. Unfortunately, the latter is not available, but the HLM requires both weights. Therefore, all the rho values are derived from the unweighted HLMs.

Appendix G: Explanatory variables used for the multivariate analysis

Variable	Description
Student-Level Variables	
SES	Pupil SES: Using the 10 possession items[31], and then the principal components analysis (PCA) method is applied to create the student SES, before it is converted into a standardised variable with mean and standard deviation equaling 0 and 1 respectively.
SES-squared	Student SES squared
Under11	Age: Under 11 years
Over11	Age: Older than 11 years
Female	Gender: Female
WC	Western Cape
NC	Northern Cape
FS	Free State
KZN	KwaZulu-Natal
NW	North West
GAU	Gauteng
MPU	Mpumalanga
LIM	Limpopo[32]
Speak1	Frequency of speaking language of test at home: Always or sometimes
Speak2	Frequency of speaking language of test at home: Unspecified[33]
Homework1	Frequency of homework given: More than once a week
Homework2	Frequency of homework given: Less than once a week
Homework3	Frequency of homework given: Unspecified[34]
NoHelp	The student receives no help with homework because he/she feels that it is unnecessary.
HelpParent	Receiving help with homework from parent or grandparent[35]
Attitude1	Index of attitude towards reading[36]: High
Attitude2	Index of attitude towards reading: Medium
Attitude3	Index of attitude towards reading: Unspecified
Safety1	Index of student safety at school[37]: High
Safety2	Index of student safety at school: Medium
Safety3	Index of student safety at school: Unspecified

[31] As mentioned in Section 3, it is assumed that the student does not possess the asset if his answer on the relevant question is unspecified.
[32] The reference dummy for the provinces is the Eastern Cape.
[33] Since nearly a quarter of students did not specify their answers, it was decided to include this dummy in order to not lose excessive sample size in the regressions. The reference group for the ‘frequency of speaking language of test at home’ variable is ‘never’.
[34] Since nearly a quarter of students did not specify their answers, it was decided to include this dummy in order to not lose excessive sample size in the regressions. The reference group for the ‘frequency of homework given’ variable is ‘never’.
[35] The reference category for the ‘receiving help with homework’ variable is that help from a parent or grandparent is not available.
[36] This index was derived using question 14 of the home questionnaire. This question asks students how strongly they agree with six different statements relating to attitudes to reading. There are 3 categories in this index: high, medium and low. Unspecified answers are also included as a dummy. The reference category is ‘low’.
[37] This index was derived using question 17 of the student questionnaire, which asks students six questions about their personal safety at school. There are 3 categories in the derived index: high, medium and low. Unspecified answers are also included as a dummy. The reference category is ‘low’.

Appendix G (Continued)

Variable	Description
	Pupil-Level Variables
Library1	Frequency of borrowing books from school or local library: At least once a week
Library2	Frequency of borrowing books from school or local library: Less than once a week
Library3	Frequency of borrowing books from school or local library: Unspecified
Book11	Number of books at home: More than 10 books
BookU	Number of books at home: Unspecified
	Home-Level Variables
Early1	Index of early home literacy activities before the child began school: High[38]
Early2	Index of early home literacy activities before the child began school: Medium
Early3	Index of early home literacy activities before the child began school: Unspecified
PreSchool	The student attended pre-school for 3 years or more
MotherMatric	Highest educational attainment of the mother: Matric or above
FatherMatric	Highest educational attainment of the father: Matric or above
ParentMatric	Highest educational attainment of either parent: Matric or above
	Teacher-Level Variables
ClassSize	Class size
TTeaching%	Percentage of time spent on teaching in class (instead of spending time on administrative duties, maintaining discipline, etc.)
TVariety	Index of variety of activities done during the reading class (e.g., asking students to read aloud, teaching students new vocabulary, etc.) using PCA[39]
TMatric	Teacher's highest educational attainment: Matric
TPostMatric	Teacher's highest educational attainment: Post-Matric qualifications
TUnspecified	Teacher's highest educational attainment: Unspecified
TYears	Years of teaching experience by the end of 2005
TAge	Teacher's age
TFemale	Teacher's gender: Female
TCareer	Index of teacher career satisfaction[40]
	School-Level Variables
Urban	Area type: Urban
Suburban	Area type: Sub-urban
UrbanSub	Area type: Urban or sub-urban
Afrikaans	Language of test: Afrikaans
English	Language of test: English
SResource	School resource shortage index, using PCA[41]
SProblems	School problem index, using PCA[42]
SAbsent1	School absenteeism[43]: serious problem
SAbsent2	School absenteeism: moderate problem
SAbsent3	School absenteeism: minor problem
SLibrary1	Availability of school library: Yes & number of books – [5000; +∞)
SLibrary2	Availability of school library: Yes & number of books – [251; 5000)
SLibrary3	Availability of school library: Yes & number of books – [0; 250)
SSES	School mean pupil SES[44]
SSES-squared	School mean pupil SES squared
Pass%	% of students scoring at or above the national average (302) in each school

[38] This index was derived using question 2 of the home questionnaire. There are 3 categories: high, medium and low.
[39] All items in question 15 of the teacher questionnaire are included when using PCA to derive this index.
[40] This index was derived using question 41 of the teacher questionnaire.
[41] All 14 items in question 18 of the school questionnaire are included when using PCA to derive this index.
[42] All 12 items in question 23 of the school questionnaire are included when using PCA to derive this index.
[43] School absenteeism is one of the problems in question 23.
[44] The standardised school mean SES is derived as follows: First, for each of the 397 South African schools, the average of the non-standardised pupil SES is taken to calculate the school mean SES. Next, these 397 school mean SES values are converted into a standardised school mean SES with zero mean and standard deviation of one.

Appendix H: Regressions on the standardised reading score

Dependent variable: Student Reading Score										
	[A]	**[B]**	**[C]**	**[D]**	**[E]**	**[F]**	**[G]**	**[H]**	**[I]**	**[J]**
Student-Level Variables										
SES	0.5169 [13.26]**	0.5597 [15.21]**	0.4949 [15.30]**	0.4440 [12.54]**	0.4113 [12.38]**	0.3831 [12.58]**	0.3671 [12.05]**	0.3868 [12.80]**	0.3707 [12.23]**	0.3400 [12.20]**
SES-squared		0.2166 [8.46]**	0.1879 [8.13]**	0.1821 [7.98]**	0.1997 [9.08]**	0.1877 [9.22]**	0.1878 [9.39]**	0.1850 [9.21]**	0.1855 [9.38]**	0.1757 [9.40]**
Under11			-0.5576 [17.04]**	-0.4680 [14.34]**	-0.4525 [14.26]**	-0.4129 [13.67]**	-0.3897 [12.99]**	-0.4147 [13.87]**	-0.3919 [13.15]**	-0.3751 [13.09]**
Over11			-0.6361 [12.07]**	-0.5340 [9.51]**	-0.5205 [9.40]**	-0.4834 [8.95]**	-0.4697 [8.64]**	-0.4762 [8.89]**	-0.4641 [8.57]**	-0.4595 [8.73]**
Female			0.2098 [7.93]**	0.2253 [9.16]**	0.2201 [9.26]**	0.2062 [8.98]**	0.1676 [7.79]**	0.2001 [8.80]**	0.1637 [7.64]**	0.1615 [7.73]**
WC				0.7858 [4.42]**	0.7607 [4.52]**	0.7220 [4.50]**	0.7145 [4.51]**	0.7261 [4.54]**	0.7183 [4.54]**	0.7022 [4.58]**
NC				0.5440 [3.23]**	0.5302 [3.36]**	0.5484 [3.70]**	0.5548 [3.80]**	0.5611 [3.83]**	0.5653 [3.90]**	0.5521 [3.90]**
FS				0.3565 [3.21]**	0.4745 [4.58]**	0.4709 [4.90]**	0.4441 [4.76]**	0.4586 [4.79]**	0.4345 [4.66]**	0.4190 [4.56]**
KZN				0.4390 [3.59]**	0.3827 [3.20]**	0.3853 [3.36]**	0.3758 [3.33]**	0.3947 [3.46]**	0.3841 [3.41]**	0.3724 [3.41]**
NW				0.2516 [1.72]	0.2455 [1.78]	0.2580 [1.97]*	0.2311 [1.80]	0.2709 [2.08]*	0.2429 [1.90]	0.2308 [1.86]
GAU				0.4715 [2.98]**	0.4787 [3.22]**	0.4623 [3.31]**	0.4540 [3.25]**	0.4811 [3.45]**	0.4702 [3.37]**	0.4564 [3.34]**
MPU				0.0173 [0.13]	0.0301 [0.25]	0.0443 [0.40]	0.0484 [0.45]	0.0605 [0.54]	0.0620 [0.57]	0.0385 [0.36]
LIM				-0.0311 [0.29]	-0.0054 [0.05]	0.0406 [0.44]	0.0441 [0.49]	0.0471 [0.52]	0.0495 [0.55]	0.0403 [0.46]
Speak1					0.4364 [13.17]**	0.3345 [11.19]**	0.3146 [10.50]**	0.3189 [10.66]**	0.3020 [10.04]**	0.2614 [9.20]**
Homework1						0.2064 [5.47]**	0.2641 [4.57]**	0.2009 [5.34]**	0.1625 [4.35]**	0.1401 [3.73]**
Homework2						0.2779 [4.85]**	0.1658 [4.43]**	0.2687 [4.76]**	0.2568 [4.49]**	0.2366 [4.19]**
NoHelp						0.3817 [8.51]**	0.3540 [8.31]**	0.3778 [8.50]**	0.3516 [8.27]**	0.3291 [8.13]**
HelpParent						0.2688 [7.34]**	0.2558 [7.20]**	0.2658 [7.43]**	0.2537 [7.27]**	0.2381 [7.08]**
Attitude1							0.3699 [10.94]**		0.3579 [10.79]**	0.3457 [10.77]**
Safety1								0.2020 [7.07]**	0.1708 [6.15]**	0.1655 [6.02]**
Library1										0.1149 [3.25]**
Book11										0.1995 [5.10]**
Constant	0.0695 [1.71]	-0.1567 [4.79]**	0.0453 [1.10]	-0.2737 [2.77]**	-0.6095 [6.95]**	-0.7737 [9.81]**	-0.8356 [10.64]**	-0.8023 [10.15]**	-0.8577 [10.89]**	-0.9032 [12.01]**
Sample size	14 657	14 657	14 343	14 343	14 343	14 343	14 343	14 343	14 343	14 343
R-squared	0.2223	0.2647	0.3409	0.3920	0.4196	0.4480	0.4703	0.4533	0.4741	0.4828

* Significant at 5%, ** Significant at 1%
Absolute values of t-statistics in brackets.

Appendix H (Continued)

Dependent variable: Student Reading Score						
	[K]	[L]	[M]	[N]	[O]	[P]
Student-Level Variables						
SES	0.3319 [12.28]**	0.3062 [12.50]**	0.3035 [12.67]**	0.2948 [12.63]**	0.3115 [12.46]**	0.3020 [11.47]**
SES-squared	0.1724 [9.43]**	0.1528 [9.22]**	0.1543 [9.31]**	0.1468 [9.14]**	0.1589 [9.29]**	0.1307 [7.93]**
Under11	-0.3705 [13.10]**	-0.3363 [12.78]**	-0.3274 [12.67]**	-0.3173 [12.49]**	-0.3357 [12.70]**	-0.3090 [11.01]**
Over11	-0.4613 [8.79]**	-0.4613 [9.30]**	-0.4626 [9.41]**	-0.4622 [9.58]**	-0.4649 [9.37]**	-0.4708 [7.96]**
Female	0.1550 [7.49]**	0.1644 [8.21]**	0.1587 [7.93]**	0.1637 [8.27]**	0.1581 [7.84]**	0.1605 [7.23]**
WC	0.7069 [4.60]**	0.6943 [4.80]**	0.7040 [4.88]**	0.6968 [4.93]**	0.6982 [4.77]**	0.5218 [3.38]**
NC	0.5574 [3.95]**	0.5485 [4.10]**	0.5664 [4.21]**	0.5584 [4.23]**	0.5561 [4.10]**	0.4578 [3.01]**
FS	0.4243 [4.57]**	0.4329 [4.83]**	0.4498 [5.11]**	0.4485 [5.13]**	0.4356 [4.86]**	0.3553 [3.63]**
KZN	0.3858 [3.50]**	0.3618 [3.47]**	0.3766 [3.67]**	0.3639 [3.60]**	0.3720 [3.55]**	0.3463 [3.17]**
NW	0.2311 [1.86]	0.2235 [1.88]	0.2162 [1.83]	0.2154 [1.85]	0.2169 [1.81]	0.2571 [1.97]*
GAU	0.4597 [3.37]**	0.4288 [3.31]**	0.4312 [3.35]**	0.4193 [3.31]**	0.4333 [3.30]**	0.3855 [3.00]**
MPU	0.0479 [0.45]	0.0484 [0.48]	0.0621 [0.62]	0.0586 [0.60]	0.0482 [0.47]	0.0651 [0.58]
LIM	0.0417 [0.47]	0.0311 [0.36]	0.0495 [0.59]	0.0407 [0.49]	0.0393 [0.46]	0.0528 [0.57]
Speak1	0.2591 [9.25]**	0.2432 [9.19]**	0.2477 [9.36]**	0.2407 [9.26]**	0.2464 [9.15]**	0.2396 [8.37]**
Homework1	0.1426 [3.87]**	0.1376 [3.95]**	0.1347 [3.86]**	0.1336 [3.92]**	0.1368 [3.85]**	0.1241 [3.23]**
Homework2	0.2378 [4.25]**	0.2327 [4.32]**	0.2289 [4.21]**	0.2280 [4.27]**	0.2338 [4.29]**	0.1951 [3.34]**
NoHelp	0.3264 [8.13]**	0.3003 [7.85]**	0.2973 [7.84]**	0.2886 [7.72]**	0.3045 [7.88]**	0.2909 [7.27]**
HelpParent	0.2336 [6.92]**	0.2048 [6.29]**	0.2017 [6.37]**	0.1920 [6.10]**	0.2043 [6.31]**	0.1858 [5.35]**
Attitude1	0.3391 [10.50]**	0.3281 [10.60]**	0.3278 [10.70]**	0.3239 [10.71]**	0.3277 [10.46]**	0.3248 [10.39]**
Safety1	0.1631 [6.03]**	0.1544 [5.80]**	0.1571 [5.91]**	0.1532 [5.79]**	0.1587 [5.88]**	0.1640 [6.41]**
Library1	0.1137 [3.25]**	0.1104 [3.35]**	0.1089 [3.32]**	0.1081 [3.37]**	0.1102 [3.27]**	0.1133 [3.23]**
Book11	0.1969 [5.07]**	0.1688 [4.66]**	0.1726 [4.87]**	0.1614 [4.63]**	0.1764 [4.85]**	0.1476 [3.57]**
Home-Level Variables						
PreSchool	0.0682 [2.19]*	0.0336 [1.20]	0.0316 [1.14]	0.0195 [0.73]	0.0352 [1.23]	0.0460 [1.55]
Early1	0.1054 [4.27]**	0.0729 [3.16]**	0.0687 [2.95]**	0.0580 [2.53]*	0.0734 [3.07]**	0.0671 [2.64]**
FatherMatric		0.4005 [11.22]**		0.2533 [8.90]**		
MotherMatric			0.4276 [12.08]**	0.3131 [10.76]**		
ParentMatric					0.3362 [11.63]**	0.3169 [10.57]**
Teacher-Level Variables						
ClassSize						-0.0096 [4.41]**
TPostMatric						0.1684 [2.87]**
TTeaching%						0.0073 [3.45]**
Constant	-0.9586 [12.54]**	-0.9867 [13.49]**	-1.0115 [14.00]**	-1.0151 [14.22]**	-1.0179 [14.08]**	-1.0256 [5.57]**
Sample size	14 343	14 343	14 343	14 343	14 343	11 942
R-squared	0.4858	0.5074	0.5114	0.5182	0.5042	0.5332

. **Appendix H (Continued)**

Dependent variable: Student Reading Score										
	[Q]	**[R]**	**[S]**	**[T]**	**[U]**	**[V]**	**[W]**	**[X]**	**[Y]**	**[Z]**
Student-Level Variables										
SES	0.2311 [11.26]**	0.2186 [10.98]**	0.2183 [11.03]**	0.2198 [10.18]**	0.2029 [10.36]**	0.2010 [10.05]**	0.1772 [8.72]**	0.1750 [7.73]**	0.0693 [5.72]**	0.0687 [5.80]*
SES-squared	0.1078 [6.79]**	0.1077 [6.90]**	0.1072 [6.89]**	0.1108 [6.79]**	0.1061 [6.69]**	0.1076 [6.78]**	0.0794 [4.62]**	0.0795 [4.80]**	0.0292 [2.63]**	0.0185 [1.91]
Under11	-0.2390 [9.48]**	-0.2272 [8.87]**	-0.2267 [8.89]**	-0.2337 [8.84]**	-0.2332 [7.48]**	-0.2242 [8.68]**	-0.2213 [7.37]**	-0.2150 [8.43]**	-0.2033 [8.43]**	-0.1788 [9.41]**
Over11	-0.3771 [7.65]**	-0.3560 [7.08]**	-0.3559 [7.03]**	-0.3808 [6.96]**	-0.3519 [5.14]**	-0.3377 [6.60]**	-0.3452 [5.38]**	-0.3268 [6.39]**	-0.3151 [7.14]**	-0.1991 [5.79]**
Female	0.1730 [8.43]**	0.1785 [8.96]**	0.1783 [8.85]**	0.1836 [8.53]**	0.1995 [9.16]**	0.1888 [9.09]**	0.1988 [9.34]**	0.1915 [9.77]**	0.1964 [11.07]**	0.1987 [12.10]**
WC	0.2127 [1.46]	0.1767 [1.18]	0.1792 [1.19]	0.1013 [0.59]	0.1118 [0.77]	0.1459 [0.93]	0.1158 [1.01]	0.1408 [1.14]	0.1058 [1.08]	0.0018 [0.03]
NC	-0.0362 [0.21]	-0.0567 [0.33]	-0.0361 [0.21]	-0.0671 [0.37]	-0.1631 [0.89]	-0.0571 [0.31]	-0.1582 [0.99]	-0.0558 [0.34]	-0.0538 [0.42]	-0.0513 [0.81]
FS	0.2635 [2.31]*	0.1992 [1.62]	0.2055 [1.67]	0.2050 [1.57]	0.1593 [1.06]	0.2004 [1.47]	0.2598 [1.95]	0.2686 [2.21]*	0.2958 [2.78]**	0.1300 [2.25]*
KZN	0.3005 [3.13]**	0.3210 [3.52]**	0.3178 [3.46]**	0.3011 [3.06]**	0.2941 [2.84]**	0.3049 [3.16]**	0.2868 [3.05]**	0.2800 [3.09]**	0.3515 [4.44]**	0.1315 [3.03]**
NW	0.3141 [2.72]**	0.2570 [2.16]*	0.2574 [2.15]*	0.3198 [2.52]*	0.3535 [2.52]*	0.3174 [2.61]**	0.3352 [2.38]*	0.3268 [2.74]**	0.2920 [2.63]**	0.0145 [0.27]
GAU	0.2980 [2.77]**	0.1604 [1.10]	0.1953 [1.54]	0.1932 [1.46]	0.1689 [1.26]	0.1650 [1.26]	0.0393 [0.33]	0.0649 [0.56]	0.1567 [1.67]	0.0525 [0.93]
MPU	0.0854 [0.74]	0.0877 [0.78]	0.0842 [0.73]	0.0478 [0.38]	-0.1454 [0.95]	-0.0097 [0.07]	-0.1119 [0.73]	0.0077 [0.06]	0.0197 [0.19]	0.0795 [1.66]
LIM	0.0910 [1.00]	0.1315 [1.51]	0.1301 [1.49]	0.0857 [0.87]	-0.0627 [0.55]	0.0116 [0.11]	-0.0634 [0.59]	0.0372 [0.38]	0.0825 [0.94]	0.0586 [1.38]
Speak1	0.2190 [7.43]**	0.2249 [7.59]**	0.2235 [7.53]**	0.2143 [6.91]**	0.2150 [5.62]**	0.2050 [6.71]**	0.1994 [5.83]**	0.1923 [6.69]**	0.1505 [5.75]**	0.1210 [6.31]**
Homework1	0.1270 [3.72]**	0.1194 [3.43]**	0.1182 [3.38]**	0.1266 [3.50]**	0.1163 [2.91]**	0.1240 [3.56]**	0.1144 [2.76]**	0.1215 [3.31]**	0.1271 [4.43]**	0.1237 [5.22]**
Homework2	0.1494 [2.89]**	0.1400 [2.76]**	0.1354 [2.62]**	0.1328 [2.39]*	0.1132 [2.06]*	0.1309 [2.49]*	0.1036 [1.78]	0.1149 [2.06]*	0.1089 [2.47]*	0.0838 [2.22]*
NoHelp	0.2224 [5.98]**	0.2107 [5.99]**	0.2106 [5.94]**	0.2107 [5.73]**	0.2157 [5.57]**	0.2013 [5.63]**	0.1980 [5.13]**	0.1837 [5.23]**	0.1533 [5.10]**	0.1397 [5.15]**
HelpParent	0.1670 [5.50]**	0.1464 [5.00]**	0.1477 [5.02]**	0.1458 [4.85]**	0.1529 [4.62]**	0.1367 [4.76]**	0.1271 [3.99]**	0.1175 [4.22]**	0.0879 [3.42]**	0.0584 [2.55]*
Attitude1	0.3269 [11.61]**	0.3123 [11.07]**	0.3150 [11.10]**	0.3001 [9.88]**	0.2897 [8.51]**	0.2951 [9.99]**	0.2785 [8.37]**	0.2891 [9.97]**	0.2971 [11.74]**	0.2634 [12.18]**
Safety1	0.1866 [7.51]**	0.1893 [7.59]**	0.1893 [7.54]**	0.1894 [7.19]**	0.1629 [5.52]**	0.1843 [7.11]**	0.1471 [5.37]**	0.1645 [6.92]**	0.1466 [6.13]**	0.0992 [4.87]**
Library1	0.0911 [3.01]**	0.0802 [2.67]**	0.0809 [2.68]**	0.0808 [2.56]*	0.0412 [1.25]	0.0692 [2.30]*	0.0229 [0.68]	0.0557 [1.76]	0.0566 [2.12]*	0.0444 [2.03]*
Book11	0.1177 [3.42]**	0.1055 [3.04]**	0.1057 [3.05]**	0.1197 [3.16]**	0.1221 [3.76]**	0.1093 [3.07]**	0.1029 [3.24]**	0.0741 [2.03]*	0.0355 [1.21]	0.0379 [2.00]*
Home-Level Variables										
PreSchool	0.0532 [1.96]	0.0563 [2.03]*	0.0554 [2.00]*	0.0586 [2.09]*	0.0544 [1.85]	0.0552 [2.08]*	0.0457 [1.43]	0.0454 [1.55]	0.0087 [0.39]	0.0199 [1.03]
Early1	0.0893 [4.02]**	0.0891 [4.03]**	0.0881 [3.95]**	0.0921 [3.89]**	0.1133 [4.84]**	0.0928 [4.01]**	0.1087 [4.91]**	0.0895 [4.14]**	0.0706 [3.60]**	0.0691 [3.83]**
ParentMatric	0.2777 [10.42]**	0.2714 [10.00]**	0.2717 [10.00]**	0.2624 [9.03]**	0.2397 [8.53]**	0.2377 [8.53]**	0.2088 [7.40]**	0.2106 [7.69]**	0.1579 [7.23]**	0.1327 [6.97]**
Teacher-Level Variables										
ClassSize	-0.0066 [3.28]**	-0.0069 [3.59]**	-0.0069 [3.55]**	-0.0065 [3.38]**	-0.0063 [3.01]**	-0.0068 [3.87]**	-0.0056 [2.81]**	-0.0062 [3.62]**	-0.0034 [2.22]*	0.0001 [0.11]
TPostMatric	0.1596 [2.44]*	0.1153 [1.72]	0.1138 [1.68]	0.1279 [1.64]	0.2161 [2.40]*	0.1127 [1.49]	0.1859 [2.19]*	0.0704 [0.99]	0.0597 [0.96]	-0.0136 [0.47]
TTeaching%	0.0022 [1.04]	0.0022 [1.06]	0.0022 [1.03]	0.0027 [1.19]	0.0039 [1.64]	0.0025 [1.14]	0.0029 [1.61]	0.0012 [0.66]	0.0002 [0.11]	0.0003 [0.32]
School-Level Variables										
Afrikaans	0.8867 [7.05]**	0.9087 [6.84]**	0.8945 [6.83]**	0.9384 [6.73]**	0.9534 [8.08]**	0.8703 [6.43]**	0.7548 [7.25]**	0.7267 [5.87]**	0.5934 [6.11]**	0.1845 [2.92]**
English	0.6599 [6.08]**	0.6161 [5.45]**	0.6110 [5.48]**	0.6256 [5.34]**	0.5910 [5.06]**	0.6270 [5.49]**	0.3199 [2.90]**	0.3919 [3.43]**	0.1604 [1.86]	-0.0361 [0.90]
Urban		0.2865 [2.51]*								
Suburban		0.2024 [2.36]*								
UrbanSub			0.2297 [2.91]**	0.1803 [2.11]*	0.2457 [2.57]*	0.1725 [1.95]	0.2327 [2.88]**	0.1931 [2.46]*	0.0254 [0.33]	-0.0729 [1.59]

SResource				-0.0294 [2.46]*	-0.0345 [2.69]**	-0.0301 [2.58]*	-0.0211 [1.60]	-0.0179 [1.51]	-0.0049 [0.47]	-0.0136 [2.25]*
SProblems					-0.0607 [4.73]**		-0.0410 [3.32]**			
SAbsent1						-0.3313 [4.17]**		-0.2494 [3.26]**	-0.1204 [1.79]	0.0147 [0.41]
SAbsent2						-0.1911 [3.19]**		-0.1366 [2.29]*	-0.0612 [1.10]	-0.0096 [0.33]
SLibrary1							0.8959 [5.93]**	0.8995 [5.94]**	0.4482 [3.33]**	0.2102 [1.99]*
SLibrary2							0.3088 [3.23]**	0.2371 [2.35]*	0.0694 [0.84]	-0.0415 [1.04]
SSES									0.3623 [8.14]**	0.1687 [5.85]**
SSES_squared									0.1314 [7.63]**	0.1048 [7.86]**
Pass%[45]										1.6084 [23.96]**
Constant	-1.0498 [6.25]**	-1.0529 [6.43]**	-1.0468 [6.29]**	-1.0703 [6.02]**	-1.1644 [6.20]**	-0.8615 [4.74]**	-1.0811 [6.97]**	-0.7918 [5.03]**	-0.7401 [5.49]**	-1.1961 [15.55]**
Sample size	11658	11658	11658	10387	8127	10286	8110	10245	10245	10185
R-squared	0.5869	0.5927	0.5923	0.5919	0.6409	0.6067	0.6641	0.6324	0.6723	0.7276

* Significant at 5%, ** Significant at 1%
Absolute values of t-statistics in brackets.

[45] Only schools with at least 15 students are included in regression [Z].