A question of efficiency: decomposing South African reading test scores using PIRLS 2006

DEBRA SHEPHERD

Stellenbosch Economic Working Papers: 20/13

KEYWORDS: BIMODAL PERFORMANCE, EFFECTIVENESS, DECOMPOSITION, SOUTH AFRICA
JEL: C31, I21, I25

DEBRA SHEPHERD
DEPARTMENT OF ECONOMICS
UNIVERSITY OF STELLENBOSCH AND
SECTION DEVELOPMENT ECONOMICS
VRIJE UNIVERSITEIT
NETHERLANDS
EMAIL: DEBRASHEPHERD@SUN.AC.ZA

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

A question of efficiency: decomposing South African reading test scores using PIRLS 2006[1]

DEBRA SHEPHERD

ABSTRACT

The "bimodal" pattern of performance observed in South Africa illustrates the persistence with which learners of former Black schools continue to lag behind their "advantaged" counterparts. It is posited that the poor functioning of former Black schools accounts for this result. A nationally representative dataset of grade 5 learners and counterfactual distribution and decomposition techniques are adopted to identify the part of the performance gap that may be explained by differences in (i) the returns structure and (ii) school characteristics composition. The former is found to be 18.9 percent of the average gap and increases significantly over the outcome distribution.

Keywords: bimodal performance, effectiveness, decomposition, South Africa
JEL codes: C31, I21, I25

[1] The author would like to thank Servaas van der Berg and Chris Elbers, as well as the colleagues in RESEP group (University of Stellenbosch) and FEWEB (Vrije University Amsterdam) for their invaluable and useful feedback.

1. INTRODUCTION

Under the apartheid government, resources for Black schools, barring "independent homeland" schools,[2] were centrally controlled by a Department of Education and Training (DET), with the control of white, Indian and coloured schools assigned to separate bodies.[3] This system led to the creation of a highly inequitable distribution of school resources across both racial and regional lines, resulting in large discrepancies in the educational attainment and performance of the different education systems. Despite concerted efforts to equalise the distribution of school resources in the South African education system, a large portion of the system, primarily historically Black schools, still fails to provide quality basic education (Van der Berg, Wood & Roux, 2002: 305). This is confirmed by the weak performance of South African students on international tests, even when compared to countries with comparatively resource-poor education systems.

Research indicates that the problem lies in the dismal performance of the historically Black school system that has failed to improve educational outcomes among the poor (Van der Berg et al, 2002; van der Berg, 2007, 2008). Recent studies have made divergent conclusions. In a cluster fixed effects analysis of schooling attainment using the first wave of the National Income Dynamics Study (NIDS) dataset, Timaeus et al (2012) argue that the poor attainment and low matriculation success of disadvantaged, mostly black, learners is not due to the poor performance of the former Black school system, but rather can be accounted for by home/parent background and socio-economic status. Although the link between race and performance is strong, black children from better socio-economic backgrounds perform exceedingly better than their less-affluent counterparts. Following democratization, there has been a "flight" of more affluent black students out of historically Black schools, with little if any movement in the opposite direction (Soudien, 2004: 104).[4] Consequently, Black schools are left with the poorest members of the community (Soudien, 2004: 106). Socioeconomic class has largely replaced race as the major determining factor of the social character or culture of a school. This may have impacts on the educational performance of historically Black schools, as the disadvantages faced by those from less affluent backgrounds are perpetuated through peer effects and low quality education. Many mechanisms exist that prevent poor children from attending affluent schools.[5]

Although racial gaps in educational attainment have been relatively closed (c.f. Van der Berg, 2008), quantitative analyses of the determinants of schooling outcomes in South Africa tends to indicate limited progress in producing equitable schooling performance for learners in South Africa. The "bimodal" pattern of test results that is typically observed illustrates how far historically Black schools continue to lag behind white, Indian and coloured schools in performance and that different data generating processes exist for historically white schools than for historically Black schools (c.f. Gustafsson, 2005; Fleisch, 2008; Van der Berg, et al., 2011; Taylor, 2011; Spaull, 2012). Indicators of school quality and

---

[2] Independent homeland schools were provided more freedom over their budgets through the South African Department of Foreign Affairs (Case & Deaton, 1999: 1049).

[3] House of Assembly for white schools, House of Representatives for Indian schools and House of Delegates for coloured schools.

[4] An example of this is provided in an article by Woolman and Fleisch (2006). They describe how Sandown High in Sandton, Gauteng, is oversubscribed whereas on the other side of town in Orlando High, Soweto, classrooms stand empty. Many of the students attending Sandown High reside close to Soweto in the Alexandra township, yet they choose to travel many kilometers to attend school elsewhere.

[5] For example, many of the affluent schools in South Africa charge fees to cover the costs of schooling not borne by the state. This power to charge fees creates an incentive to admit as many full fee-paying students as the school can accommodate (Woolman & Fleisch, 2006: 32).

resource allocation such as teacher-pupil ratios and school fees charged often fail to explain the poor performance of Black and Coloured schools (c.f. Van der Berg and Burger, 2003). This may be due to the fact that the conventional measures of school quality may not in fact be correct measures for explaining the impact of school-related factors in education production functions. For example, the impact of smaller classroom sizes and lower pupil-teacher ratios may appear insignificant not because school quality does not matter for educational achievement, but rather because impact of school quality works through other, generally immeasurable, mechanisms which are either directly or indirectly related to classroom size and pupil-teacher ratios such as motivation of teachers and managerial capacity of the principal. Home background factors such as socio-economic status and parent education are found to be significant in explaining the variation in performance results (c.f. Taylor and Yu, 2008; Van der Berg, 2008) and attainment (c.f. Timaeus and Boler, 2007; Lam et al, 2011; Timaeus et al, 2012). However, these variables are most likely positively related to unobservable home background characteristics that are themselves related to school choice such as the value that parents place on education.

This paper aims to shed light on the source/s of discrepancy in performance between former Black and former advantaged schools, and whether the discrepancy comes as a result of differences in school quality[6] or access to a lower level of (quality) resources. We employ mean and quantile regression decomposition techniques to derive a counterfactual outcome distribution that allows us to quantify the distributional effects of educational input endowment (explained effect) versus the returns to these inputs (unexplained effect). The part of the gap left unexplained has been relatively ignored in the literature. This is partly due to an emphasis being placed on comparing the relative importance of levels of educational inputs at the level of the home and school and partly due to the ambiguity surrounding the correct interpretation of the unexplained gap. A substantial unexplained component would indicate a stronger ability of one group of students, or school type, to transform schooling inputs into outputs. In other words, the unexplained component provides an indication of the difference in quality between schools. This study aims to focus more attention to the unexplained component.

The remainder of the paper is laid out as follows: Section 2 introduces the data and summary statistics for the two school groups under comparison. Section 3 presents the empirical model and methodology including extensions and issues and section 4 discusses the results. Section 5 concludes.

2. METHODOLOGY

2.1 Oaxaca-Blinder decomposition

Decomposition methods, starting with the seminal work of Oaxaca (1973) and Blinder (1973) and the so-named Oaxaca-Blinder (OB) decomposition, finds its roots in the labour market literature. Its adoption in the context of educational outcomes is fairly recent, where studies have chiefly emanated from the education production function literature in an attempt to explain the extent to which performance gaps may be explained by differences in learner and school characteristics, with the remaining gap due to differences in the quality or effectiveness of the different education processes. Applications exist across geographical lines (c.f. Tansel, 1998; McEwan, 2004; Ammermuller, 2007; Burger, 2011), school types (c.f.

[6] School quality is defined as the extent to which a school and its constituent parts (teachers, management, culture and infrastructure) improve a student's learning.

Krieg and Storer, 2006; Duncan and Sandy, 2007), across time (c.f. Barrera-Osorio et al, 2011; Cattaneo and Wolter, 2012; Sakellariou, 2012; Da Maia, 2012) and across race and gender (Sohn, 2008; 2010).

The last two decades have witnessed extensions and improvements to the decomposition methodology, including an extension of the OB methodology to the entire distribution of, for example, wages and test performance, as well as links made to the treatment effect literature. Recent contributions by Barsky et al (2002), Fortin et al (2011) and Sloczynski (2012) have shown that the Oaxaca-Blinder decomposition provides a consistent estimator of the population average treatment effect of the treated (PATT). Kline (2011) has further shown the method to be equivalent to a propensity score reweighting estimator that is based on a linear model for the odds of being treated and provides a "doubly robust" estimator of the counterfactual mean.[7]

We consider a population of N learners indexed by $i = 1, \ldots, N$ that are divided into two mutually exclusive groups denoted by the binary variable g_i where $g_i = 0$ represents membership to the group of historically disadvantaged schools (control group) and $g_i = 1$ represents membership to the group of historically advantaged schools (treatment group). The outcome of interest is the reading test score Y_{ig}. We further observe a set of k controls X_i. As in the treatment effect literature, Y_{i0} and Y_{i1} can be interpreted as two potential outcomes for learner i. Although both of these outcomes are observed, only one is realized, with the realized outcome given by:

$$Y_{ig} = Y_i(0)(g_i - 1) + Y_i(1)g_i \tag{1}$$

The Oaxaca-Blinder model is based on a model for the potential outcomes that is linear and allows the regression coefficients across the two groups to be different:

$$Y_{ig} = X_i'\beta_g + \varepsilon_{ig} \quad \text{where} \quad E[\varepsilon_{ig}|X_i, g_i] = 0 \quad \text{for} \quad g\epsilon\{0,1\} \tag{2}$$

There are three possible reasons why the distribution of reading scores between the two school types could differ: i) differences between the returns structures β_0 and β_1; ii) differences in the distribution of observable characteristics X; and iii) differences in the distribution of unobservable characteristics ε. The aim of decomposition is to separate the contribution of (i) from (ii) and (iii).[8] Knowledge of β_0 and β_1 allows us to compute a counterfactual of the type "what would be the distribution of reading scores for learners in group 0?", and vice versa. A counterfactual distribution of this type allows us to decompose differences in the performance of learners in school type 0 and those in school type 1 into a component attributable to differences in the observed characteristics of learners and their schools (explained component) and a component attributable to differences in the returns structure to these characteristics (unexplained component). The mean reading score gap may be decomposed as follows:

[7] If the true offs of treatment is linear in then the Oaxaca Blinder estimate of the average treatment effect will be identified even if the model for potential outcomes is misspecified, provided that unconfoundedness and overlap hold. Conversely, if the model for potential outcomes is correct, the Oaxaca Blinder estimate will identify the average treatment effect even if overlap fails and/or the implicit model for the odds of treatment is incorrect.

[8] In order for the decomposition to follow a partial equilibrium approach, we restrict the counterfactual returns structure to one of a "simple" counterfactual treatment in that the only alternative state of the world for group A would be the returns structure faced by group B, and vice versa. This assumption rules out the existence of some other counterfactual returns structure that would prevail if, for instance, learners from advantaged schools were no longer enrolled in those schools.

$$E[Y_i|g_i = 1] - E[Y_i|g_i = 0] = E[X_i|g_i = 1]\beta_1 - E[X_i|g_i = 0]\beta_0$$
$$= E[X_i|g_i = 1]\beta_1 - E[X_i|g_i = 0]\beta_1 + E[X_i|g_i = 0]\beta_1 - E[X_i|g_i = 0]\beta_0$$
$$= (E[X_i|g_i = 1] - E[X_i|g_i = 0])\beta_1 + E[X_i|g_i = 0](\beta_1 - \beta_0) \quad (3)$$

The first term in the final line of (3) represents the explained component of the performance gap that is due to differences in the average endowments of individual and/or school resource/quality characteristics. The second term represents the "discrimination" or unexplained component of the wage gap.[9] From Sloczynski (2012), the unexplained component of the Oaxaca Blinder decomposition represents the PATT in the treatment literature:

$$E[Y_i|g_i = 1] - E[Y_i|g_i = 0] = (E[X_i|g_i = 1] - E[X_i|g_i = 0])\beta_1 + E[X_i|g_i = 0](\beta_1 - \beta_0)$$
$$= E[Y_{i1} - Y_{i0}|g_i = 1] + \{E[\mathrm{Y}_{i0}|g_i = 1] - E[\mathrm{Y}_{i0}|g_i = 0]\}$$
$$= \tau_{PATT} + \{E[\mathrm{Y}_{i0}|g_i = 1] - E[\mathrm{Y}_{i0}|g_i = 0]\} \quad (4)$$

The test score gap can therefore be decomposed into the PATT (unexplained component) and "selection bias" which represents the extent to which the control group (0) and treated group (1) are on average different (explained component).

As in the treatment literature, a number of assumptions need to be made in order to identify the PATT. The first of these is the ignorability or unconfoundedness assumption which states that the distribution of unobservable determinants of test performance are the same across both groups after controlling for observable characteristics; that is, $g_i \perp Y_{i0}, Y_{i1}|X_i$. This rules our selection into group 1 or 0 based on unobservables. Secondly, we assume that there do not exist any (sets of) values of X which would perfectly predict membership to either group 0 or 1; that is, $\Pr(g_i|X_i = x) < 1$ for all x. This is known as overlapping support assumption. Finally, we make an assumption of simple treatment counterfactuals (no general equilibrium effects) such that the coefficients of the education production process at type 1 schools would remain unaltered if the composition of the school and it's student body were to take on the average characteristics of type 0 schools, and vice versa.[10] This implies that we cannot give a "truly" causal interpretation to the estimates of the Oaxaca Blinder decomposition.

Following estimation of the aggregate decomposition we may further investigate the individual and collective contributions of individual characteristics to each of the decomposition components. It is fairly simple to identify the contributions of individual characteristics to the explained component given that the total component is merely a sum over the individual contributions (Jann, 2008: 8). Interpreting a detailed decomposition of the unexplained component is less straightforward as issues arise when the explanatory variables of interest are categorical and do not have an absolute interpretation (Fortin, 2010). Specifically, two problems are generated: first, when scalable variables do not have a natural

[9] Equation (3) can also be represented in terms of simple counterfactural distributions as $E[X_i|g_i = 1]\beta_1 - E[X_i|g_i = 1]\beta_0 + E[X_i|g_i = 1]\beta_0 - E[X_i|g_i = 0]\beta_0$

[10] The choice of whether to construct the counterfactual from the returns structure of group 1 or 0 corresponds to two methods of decomposing the differences in student characteristics (Krieg & Storer, 2006: 569). The research question posed by this study favours the use of group 1 returns structure in order to calculate the counterfactual distribution as we ask the question: what if students attending historically Black schools received the same treatment as students attending historically advantaged schools, and if so, what would the gap in reading scores be?

zero implying that the reference point has to be chosen arbitrarily; and second, the choice of reference category will both change the individual contributions of single categorical variables to the unexplained component as well as alter the contribution of the category as a whole (c.f. Jones, 1983; Oaxaca & Ransom, 1999; Nielsen, 2000; Horrace & Oaxaca, 2001; Yun, 2005). A simple solution to the identification problem associated with the choice of reference category has been proposed by Yun (2005). The intuition behind the solution is to obtain the "true" contributions of individual variables to the reading score gap as the average of the regression estimates obtained from every possible specification of the reference groups. We constrain the coefficient on the omitted category to be equal to the un-weighted average of the coefficients on the other categories:

$$\beta_{g,C_1} = -\sum_{k\neq} \beta_{g,C_k}/K \text{ and } \sum_{k=1}^{K} \beta_{g,C_k} = 0 \quad (5)$$

where C is a categorical variable with K = 1, 2, …., k categories and k = 1 is the omitted group. This new set of "normalised" regression coefficients on the dummy variables and constant can be used with the original set of explanatory variables, X, to obtain the "true" contributions of different variables to the two components of the reading score gap. It is simple to show that the total explained and unexplained components given by the untransformed model in equation (2) are identical to the estimates given by the transformed model using the Yun transformed regression coefficients.[11] Standard errors for the individual contributions are furthermore straightforward to estimate (Jann, 2008).

2.2 Decomposition at percentiles

The traditional Oaxaca-Blinder decomposition estimates the unexplained and explained components only at the mean. However, we may be interested in decomposing the performance gap at various points of the outcome distribution. The labour market literature offers a number of approaches to obtain the counterfactual distribution (c.f. Juhn et al, 1993; DiNardo et al, 1996; Machado and Mata, 2005; Autor et al, 2005). One such approach involves replacing the conditional distribution for one group of the two groups of interest with that of the other group. This requires directly estimating the conditional distribution of outcomes for both groups. Two early parametric methods for doing this were suggested by Donald et al (2000), and Fortin and Lemieux (1998). A more recent quantile-regression based approach proposed by Chernozhukov et al (2012) evaluates distributional effects using conditional quantile regressions. The analysis in this paper adopts this methodology. The Chernozhukov et al (2012) approach focuses on estimating the counterfactual distribution for group 0 learners that would have prevailed had they faced the conditional distribution of group 1. Let $Q_{Y_g}(\pi|x)$ and F_{X_g} represent the conditional π-quantile of Y given X in group g and the marginal distribution of covariates in group h, for $g, h \in \{0,1\}$. The observed Y_g^g outcome can be defined as $Y_g^g = Q_{Y_g}(v_g^g|X_g)$ where $v_1^0 \sim v(0,1)$ independently of $X_g \sim F_{X_g}$. As with the aggregate decomposition, we consider a counterfactual that is computed through combining the conditional distribution in group 1 with the covariate distribution in group 0; that is, $Y_1^0 = Q_{Y_1}(v_1^0|X_0)$ where $v_1^0 \sim v(0,1)$ independently of $X_0 \sim F_{X_0}$.[12] The conditional distribution is related to the quantile function as follows:

[11] While these restrictions may appear to solve the problem of the omitted group, Yun (2008) points out that there remains a degree of arbitrariness to the methodology and caution should be taken when interpreting the detailed decomposition results.

[12] Construction of the counterfactual $Y_1^0(Y_0^1)$ assumes that $Q_{Y_1}(v|x)$ $(Q_{Y_0}(v|x))$ can be evaluated at each x in the support of X_0 (X_1), or that the quantile function can be extrapolated outside of the support of X_0 (X_1). Given that support does not hold for all covariates, particularly school average SES, we assume the latter.

$$F_{Y_g}(y|x) = \int_0^1 1\left\{Q_{Y_g}(\pi|x) \leq y\right\} d\pi \quad (6)$$

The marginal distribution and quantile function of interest are therefore:

$$F_{Y_1}^0(y) = \int F_{Y_1|X_1}(y|x) dF_{X_1}(x) \quad (7)$$

$$Q_{Y_1}^0(\pi) = \inf\left\{y: F_{Y_1}^0(y) \geq \pi\right\}, \pi \in (0,1) \quad (8)$$

The π-quantile and y-distribution treatment (unexplained) effects are determined as follows:

$$QE_{Y_1}^0(\pi) = Q_{Y_1}^0(\pi) - Q_{Y_0}^0(\pi) \quad (9)$$

$$DE_{Y_1}^0(y) = F_{Y_1}^0(y) - F_{Y_0}^0(y) \quad (10)$$

The marginal distribution and quantile functions expressed in (7) and (8) above depend on either the underlying conditional quantile function or conditional distribution function. The former is estimated using the linear quantile regression method of Koenker and Basset (1978) and Koenker (2005). Quantile regression can be a very flexible technique given a rich set of covariates that allows the true conditional quantile function to be estimated arbitrarily well, but only in the case where y has a smooth conditional density (Koenker, 2005). Chernozhukov et al (2009) suggest a more flexible distribution regression approach, adopted by this study, to estimate the conditional distribution that uses a separate regression model for each value of y in order to compute $F_Y(y|x)$.[13] Consider the following model:

$$F_Y(y|x) = \Lambda(m(y,x)) \quad (11)$$

where, for purposes of this study, Λ is a probit function. Separate probits are estimated at each value of y. The counterfactual distribution $F_{Y_1}^0(y)$ is then estimated by averaging over the predicted probabilities and inverting to obtain the counterfactual quantiles as in equation (8). An important advantage of this distribution regression approach is that it can be generalized to the case of the detailed decomposition. Specifically, the detailed decomposition of quantiles is obtained by computing the different counterfactuals for each element of X sequentially and then inverting to get the counterfactual quantiles. However, the results will be path dependent.[14]

The cdeco and counterfactual STATA command is used to estimate the conditional and counterfactual quantiles at each percentile of the test score distribution (Chernozhukov et al, 2008). Bootstrapped standard errors are generated using 100 repetitions.

2.3 Methodological issues

[13] Although this method is similar to quantile regression, it does not suffer from several issues associated with quantile regression. Specifically, quantile regression may involve substantial rounding and may not provide a good approximation to conditional quantiles where the conditional quantile function may be highly linear (Chernozhukov et al, 2009).

[14] No generalised view exists that favours using the one method over the other. In practice, the choice to use quantile regression or distribution regression models depends largely on the empirical performance of each and their ability to handle complicated data situations (Chernozhukov et al, 2012). The distribution regression approach is chosen for this study as it does not require smoothness of the conditional density and is generalizable to the detailed decomposition. Given common support over the covariates, the results of both methods will coincide. However, when common support is not fully satisfied the results can be markedly different.

Multicollinearity and endogeneity bias are prominent issues in the estimation of education production functions. Although these issues may not reduce the predictive power or reliability of the model as a whole, they potentially bias the estimates on individual predictors. A more serious problem than multicollinearity may be that of measurement error and missing data. Measurement errors tend to be most severe in the case of school inputs. For more detailed discussions of these and other problems, see Hanushek (1979), Todd and Wolpin (2003) and Webbink (2005).

Of more concern to this study is violation of the somewhat strong assumption of conditional independence made earlier. It is highly plausible that parents may select the schools which their children attend. For example, certain schools attract students from higher socio-economic backgrounds because their parents wish their children to attend the best available schools. Similarly, student sorting may result if parents choose to reside in areas where good schools are easily accessible. If this is the case, differences in student body composition would not be wholly exogenous and the conditional distribution of $(X, \varepsilon)|DB = 1$ may be different from the distribution of $(X, \varepsilon)|DB = 0$. The conditional independence assumption does not necessarily rule out the possibility that these distributions may be different, but it constrains their relationship. Specifically, the joint densities of X and ε for groups A and B have to be similar up to a ratio of conditional probabilities (Fortin et al, 2010).

The literature offers several solutions to deal with violation of the conditional mean independence assumption, the traditional methods being the use of control function[15] (Heckman, 1974, 1976; Heckman and Robb, 1986; Heckman and Hotz, 1989) or instrumental variable models (Heckman and Vytlacil, 1999, 2001, 2004; Heckman, 2001). However, it is difficult to find a credible excluded instrumental variable for the choice into a former advantaged school. Arguably the best way of dealing with selection and endogeneity is to use panel data methods. Given the cross-sectional nature of the data employed by this study, we need to be mindful of potential bias in the model parameters when interpreting the results.

3. DATA AND SUMMARY STATISTICS

The Progress in International Reading Literacy Study (PIRLS) conducted in 2005/6 by the IEA[16] was the second of its kind conducted in a five year cycle (after PIRLS 2001) in which particular emphasis was placed on the reading proficiency of young children. Although the survey collected data on 45 schooling systems from 40 countries, only the South African data is used for purposes of this paper.[17] Grade 4 learners were tested, with the exception of Luxembourg, New Zealand and South Africa, where learners were sampled from the fifth grade. In addition to the collection of reading test scores, a full array of background information regarding home and school environments was collated. 14125 grade 5 students were sampled from 385 schools in South Africa. The relatively large size of the dataset makes PIRLS 2006 highly advantageous for analysing educational outcomes and its determinants in

[15] The inclusion of a control variable such as the usual inverse Mills'ratio changes the interpretation of the decomposition (cf Fortin et al, 2010).
[16] International Association for the Evaluation of Educational Achievement.
[17] There may be concern that the developed country context of the PIRLS study may have generated a bias in the South African reading scores in favour of English speaking students in wealthier schools. However, similar performance gaps between rich and poor schools (as proxies for the former school departments) have been observed in regional studies (c.f. van der Berg, 2008; Spaull, 2011; 2012).

South Africa, as previous research has revealed a very large intra-class correlation coefficient in South Africa of around 0.7 for reading scores (see for example Van der Berg, 2006). The sample of schools needs to be suitably large such that the sample variation in schooling outcomes truly reflects that observed in the South African education system. Of all the countries that participated in the PIRLS 2006 survey, the situation in South Africa proved to be the most complex given that the questionnaires and assessment tools had to be translated into all of the 11 official languages.

As this study is interested in the observed performance gap between historically Black and historically advantaged schools, the sample of students needed to be divided into these two school types. The dataset provides no information of the former school department, but schools were able to select the language of the test. It is safe to assume that schools that tested in an African language would have fallen under the historically Black system. It is furthermore likely that schools formerly belonging to the white, Indian and coloured education departments would have tested in English or Afrikaans. However, an overlap between the two groups may exist in that a number of formerly Black schools may have tested in (particularly) English.[18] Therefore we will refrain from using the distinction of former disadvantaged and advantaged schools and rather denote the groups as English/Afrikaans testing schools and African language testing schools. In order to address the issue of overlap between the two groups, a further restriction was applied to the sample of formerly advantaged schools. If more than 65 percent of the grade 5 sample from a particular school was found to not speak the test language on a regular basis, this school was dropped from the group of English/Afrikaans testing schools. The decision to drop schools and not simply move them to the sample of African language testing schools was made as some of the schools meeting the aforementioned restriction may not in fact be historically Black schools. In fact, some of the schools may be historically coloured schools that are poor and weak performing. Consequently, the remaining sample of English/Afrikaans testing schools may suffer from positive selection bias if we assume that the remaining group of schools are the richer, and hence better performing schools. This should be kept in mind when interpreting the results.

Estimates based on the full sample of English/Afrikaans testing schools will serve as a robustness check to the main results. The dependent variable employed in the empirical model is the individual student reading score.[19] The international scores are set on a scale with a mean of 500 and a standard deviation of 100. In the process of choosing covariates to be regressed on the student reading score, two new variables had to be generated. These were a wealth measure of a pupil's household represented by a socio-economic index, and a school resource index. Both of these indices were generated using the first principal component approach.[20] Definitions of all variables included in the empirical model are provided in the appendix. The main problem posed by the data was that of a large number of missing data, particularly at the student level. Dropping these students would reduce the amount of variation in the dependent variable, causing bias in the results (Ammermuller, 2007). A brief note on the imputation methods used to deal with missing data at the household level is provided in note 1 of the appendix. Given the comparatively smaller number of missing data, schools with missing data were dropped from the sample. The final sample includes 9134

[18] In a separate study by Desai (2001), a primary school in the Khayelitsha township, Cape Town, was observed where the home language of the majority of learners and educators was Xhosa. However, since 1995 the school has decided to use English as the medium in which all school work is to be expressed from grade 4, although this does not prevent the teachers from relaying information to the students in an African language.

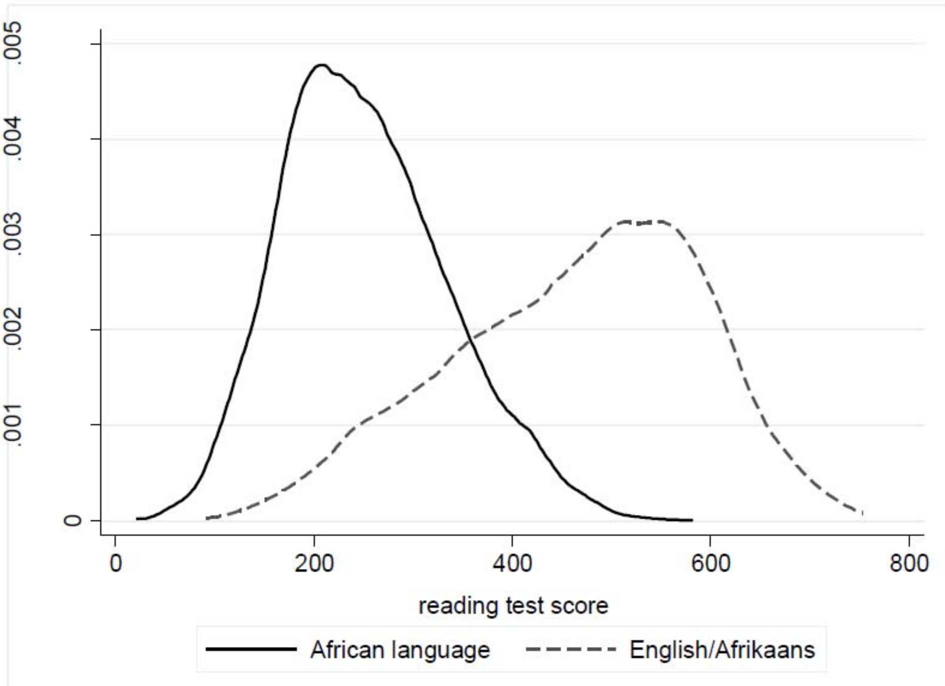
[19] The rest score is calculated using average scale scores computed from 5 plausible imputed scores based on Item Response Theory (IRT).

[20] Pearson (1901).

students in 240 African language testing schools, and 2107 students in 66 English/Afrikaans testing schools. This is similar to what is observed in the South African education system: 21 percent formerly “advantaged” schools and 79 percent formerly Black (disadvantaged) schools.[21]

The average reading scores are 252 and 465 for African language and English/Afrikaans testing schools respectively, representing a statistically significant performance gap of 213 points. This difference is dramatic when viewed in the context of 50 points on the PIRLS test being described as equivalent to one school grade (Filmer et al, 2006). South African learners attending both school types performed lower than the international average. Both a higher mean and test score spread for English/Afrikaans testing schools is depicted in Figure 1. Figure A1 of the appendix shows the distribution of African language testing schools compared to English/Afrikaans testing schools with and without sample restrictions. It is clear that the excluded schools are predominantly worse performing ones, yet even after the restrictions are made a significant proportion of learners in the group of English/Afrikaans testing schools are performing at quite low levels.[22] This is due to the fact that the former may include a number of coloured schools that may have similarly low SES levels as African language testing schools. This is particularly the case among coloured schools that are comparatively poorer than their affluent counterparts within the same school grouping.

Figure 1: reading score distributions, by school type



[21] “Advantaged” here refers to schools that did not fall under the former Black (DET and homeland) school system and therefore may include former White, Coloured and Indian schools. The former DET and homeland schools make up approximately 80 percent of South African primary schools. Some of the “advantaged” schools, particularly coloured schools, are not likely to be wealthy, well-functioning schools.

[22] Figure A2 compares the reading score performance of the two school groups under consideration in this study to the literacy test score distributions of grade 4 learners by former department from the NSES study conducted in 2008. Typically former department is proxied by school wealth (c.f. van der Berg, 2008; Taylor and Yu, 2008; Spaull, 2012). However, from figure A2 it appears that the division based on language of test proxies closer to the former white school system than using the top 20 percent wealthiest schools based on average school SES.

Table A1 of the appendix summarises descriptive statistics on all control variables included in the empirical model by school type. There are clear differences in the composition of the student body and school functioning between the two school types. Comparisons of the means indicate that learners attending English/Afrikaans testing schools are significantly less likely to be overage or underage, as well as significantly more likely to speak the test language at home on a regular basis. Furthermore, the learners are more likely to receive help with their reading homework, have better educated parents with full-time employment, come from households with higher SES and engage in more reading activities at home. African language testing schools report higher levels of absenteeism and are significantly poorer on average as measured by the average SES of the student body.

English/Afrikaans testing schools are found to have significantly greater parent involvement and generally do not provide free/subsidised lunch programmes to their learners. At the classroom and teacher levels, there is a significantly larger proportion of teachers with teaching diplomas and greater usage of higher-order reading aides in English/Afrikaans testing schools. On the other hand, a significantly larger proportion of teachers in African language testing schools report a greater variety of daily use of in-class learning and teaching activities and methods and diagnostic testing. However, this does not allude to how much time is spent on each activity, which might vary between schools. Differences in the endowments of such variables may explain the observed reading test score gap, although this inference needs to be tested using the methodology outlined in this paper.

4. EMPIRICAL RESULTS

4.1 Regression results

Regression outputs of the final model specification for each school sample are presented in table A2 of the appendix. The model fits the sample of English/Afrikaans schools quite well, as observed by an adjusted R-squared of 0.71. The adjusted R-squared is much lower in the sample of African language testing schools (0.33).[23] A number of inputs are found to be similarly related to educational performance across the two school samples. A common finding in the literature is that girls perform on average better in literacy and reading. Parent education and employment as well as teacher education and experience are positively related to test scores for both samples, with no significant difference in the estimated coefficients. In addition, the positive and significant relationship between performance and frequency of test language spoken at home and household SES is estimated to be significantly larger for the group of English/Afrikaans testing schools (see the last column of table A2).

Differences in school functioning across the two samples are evident at the school and classroom levels. Learners attending African language schools benefit significantly from classroom discussion and interaction, frequency of homework, diagnostic testing and

[23] It is interesting to note that pupil and family background characteristics (specification 1) alone explain quite a large proportion of the variation in learner test scores as reflected by the R-squared (0.56 for English/Afrikaans testing schools and 0.22 for the African language testing school sample). The addition of province controls to specification [1] (not shown here) results in an increase in the R-squared value for the African language testing sample to 0.29, while it only increases by a further percentage point after the inclusion of school controls. In the case of the English/Afrikaans testing schools, the addition of provincial dummies increases the R-squared marginally to 0.61, whilst the addition of school controls leads to a further increase to 0.69. Changes in the R-squared values once accounting for further controls may indicate variation in performance across provinces for African language testing schools which may be linked to differences in provincial school functioning. In the case of English/Afrikaans testing schools, there appears to be within province variation in performance and hence school functioning.

extended learning time. The contrary is true for the sample of English/Afrikaans testing schools. English/Afrikaans schools with high parent involvement and higher order learning aides such as reading series and long books with chapters perform better on average. School SES is similarly estimated to have a large and highly significant coefficient for the group of English/Afrikaans schools but not for the group of African language schools. Furthermore, whereas teacher age and performance do not appear to be related in English/Afrikaans schools, learners taught by younger teachers in African language schools perform significantly better than those learners taught by teachers older than 30. This may be related to changes that have taken place in teacher hiring and training practices. Specifically, the 2007 National Policy Framework for Teacher Education stipulates that primary school teachers entering the profession are required to obtain a minimum of a four-year tertiary qualification. In 2004, only 48 percent of teachers met this minimum qualification. The age distribution of teachers in PIRLS 2006 indicates that less than 5 percent of grade 6 reading teachers were younger than 30 years of age. Of these, approximately 50 percent had at least a university degree (or equivalent). This is compared to 36 percent of teachers who are 30 years or older.

4.2 Oaxaca-Blinder and detailed decomposition results

The estimated models for the two school groups are used to decompose the test score gap into the explained and unexplained components. Estimates of the explained and unexplained gaps shown in table 1 indicate that 81.1 percent of the test score gap between African language testing and Afrikaans/English testing schools can be explained by differences in average endowments of student, household and school characteristics. The remaining 18.9 percent represents the unexplained gap that is due to differences in school efficiency. Both the explained and unexplained gaps are statistically significant. The decomposition results therefore suggest that former advantaged schools and their students are both more endowed with those characteristics conducive to higher schooling outcomes and more efficient in transforming educational inputs into educational outcomes. In other words, keeping the distribution of characteristics of African language schools and their learners the same but facing the English/Afrikaans school returns to these characteristics, learners' test scores would be improved by an average of 40 points.

Table 1: Oaxaca-Blinder decomposition results, English/Afrikaans school returns as counterfactual [a, b]

Average test score:	
English/Afrikaans schools (g=1)	464.6
African language schools (g=0)	251.6
Average test score gap	213
Explained gap	172.7***
(robust standard error)	(17.93)
Proportion of average gap	81.1
Unexplained gap	40.3**
(robust standard error)	(11.44)
Proportion of average gap	18.9

[a] * p<0.10; ** p<0.05; ***p<0.01
[b] Robust clustered standard errors in parentheses.

The results of the detailed decompositions are presented in table 2. The explained and unexplained effects have been grouped according to student/household, school and classroom/teacher characteristics.[24] School SES has been kept as a separate category. It is clear that the largest portion of the performance gap between African language and English/Afrikaans testing schools is due to differences in the average endowments of school characteristics, particularly school SES. Differences in school characteristics (including school SES) account for 52 percent of the total performance gap. A further important contributing component is that of parent involvement. This is estimated to account for 11 percent of the total gap. School SES and parent involvement may be proxies for a number of institutional, organisational and cultural processes within schools. Therefore, we may posit that the poor performance by students who attended African language testing schools could largely be explained by the fact that they lack access to those individual and school characteristics which lead to better schooling outcomes.

Table 2: Detailed Oaxaca-Blinder decomposition results [a, b]

	Explained gap (robust s.e.)	Unexplained gap (robust s.e.)
Learner/household	61.38***	-17.05***
	(7.22)	(6.23)
School	4.58	22.56***
	(9.53)	(6.92)
School SES	86.04***	-19.76***
	(14.71)	(5.16)
Class/teacher	11.09*	16.23
	(6.41)	(14.10)
Province	9.62**	-2.76
	(3.99)	(5.36)
Constant	-	-41.05**
		(-17.96)
Gap totals	172.7***	40.3***
	(17.93)	(11.44)

[a] * p<0.10; ** p<0.05; ***p<0.01
[b] Robust clustered standard errors in parentheses.

The contributions of individual variables to the unexplained component indicate the efficiency with which former Black and former advantaged schools convert the respective educational inputs into better schooling outcomes. The negative sub-total on pupil/household indicates that African language testing schools were more efficient in transforming these inputs into educational outcomes. Regarding school and classroom/teacher level factors, the positive contributions to the unexplained component indicates that English/Afrikaans schools and their students are more efficient in transforming these inputs into better test scores. However, we will refrain from placing too much emphasis on the interpretation of the unexplained component given the well cited issues with attributing the unexplained effect to individual (groups of) covariates.

Returning to equation (1), the size of the explained component is dependent on two factors: the difference in the average endowments between the two school types; and the

[24] Detailed results broken down to each covariate is available on request from the author.

coefficient structure of the English/Afrikaans testing schools. The selection issues and endogeneity biases discussed previously may point towards a potential upward bias in the English/Afrikaans school model coefficients that would increase the relative size of the explained component. The upward bias may be driven by, for example, exclusion of predominantly weaker historically advantaged schools from the sample of English/Afrikaans testing schools and selection into the wealthier, better performing former advantaged school system driven by unobservable factors which are positively related to schooling inputs and processes. Given the large positive coefficients on, in particular, school SES, parent involvement and teacher education in the English/Afrikaans schools model (see table A2 of the appendix), as well as the higher average endowments in favour of these schools, it is unsurprising that the decomposition yields a large and significant explained component. School SES alone explains about 40 percent of the overall average test score gap. Correction for selection and endogeneity biases may result in different relative sizes of the explained and unexplained coefficients. If the bias in the coefficients is indeed upward, the contribution of the explained component to the overall test score gap may be smaller than is estimated by this study.

4.3 Sensitivity testing

One criticism of the analysis presented may be the large number of "unspecified" or missing data in the pupil and household covariates that were entered into the regressions in order to avoid dropping large numbers of missing data at the student level. A further concern of the analysis may be the effect of the restrictions made on the former advantaged school sample. If the impact was to mistakenly exclude poor performing former advantaged schools from the sample, this may have led to an upward bias in the estimates through sample selection effects. It would therefore be useful to test whether the aggregate decomposition results differ substantially after allowances are made for these missing data. The results are shown in table 3. The first column indicates the aggregate decomposition results from the headline specification in table 1. Column 2 and 3 indicate estimates for samples that excludes observations for which there was missing information on learner and teacher characteristics respectively. Excluding missing data on learner and teacher covariates results in excluding between 20 to 25 percent and 25 to 40 percent of learners from the English/Afrikaans and African language school groups respectively. It is clear that missing data is most prevalent amongst the weaker performing students as the average reading scores for both school groups have increased, whilst the size of the average performance gap has remained unchanged. The decomposition results are robust to the exclusion of missing data at the learner and teacher level.

Columns 4 and 5 indicate estimate results whereby the schools originally excluded from the analysis are re-assigned to one of the two schools groups. The excluded schools are first assigned to the group of English/Afrikaans testing schools (column 4) and subsequently the group of African language testing schools (column 5). As is expected, reassignment of these schools serves to lower the average performance of the English/Afrikaans testing group and increase the average performance of the African language testing group. Therefore, the average gap is decreased in the case of the former and increased in the case of the latter. Despite this, the relative sizes of the explained and unexplained component are fairly robust to the headline estimates.

Table 3: Oaxaca-Blinder decomposition results, sensitivity tests [a, b]

	1	2	3	4	5
Average test score:					
English/Afrikaans schools (g=1)	464.6	477.4	470.9	406.3	464.6
African language schools (g=0)	251.6	265.3	253.3	251.6	270.1
Average test score gap	213.0	212.1	217.6	154.7	194.5
Explained gap	172.70***	176.89***	176.6***	132.63***	157.0***
(robust standard error)	(17.93)	(18.41)	(19.29)	(15.21)	(17.37)
Proportion of average gap	81.1	83.4	81.2	85.7	80.7
Unexplained gap	40.30**	35.25***	40.99**	22.05**	37.49***
(robust standard error)	(11.44)	(11.66)	(11.88)	(9.54)	(10.14)
Proportion of average gap	18.9	16.6	18.8	14.3	19.3
Observations in g=1	2107	1563	1694	4092	2107
Observations in g=0	9134	5591	6713	9134	11119

[a] * p<0.10; ** p<0.05; ***p<0.01
[b] Robust clustered standard errors in parentheses.

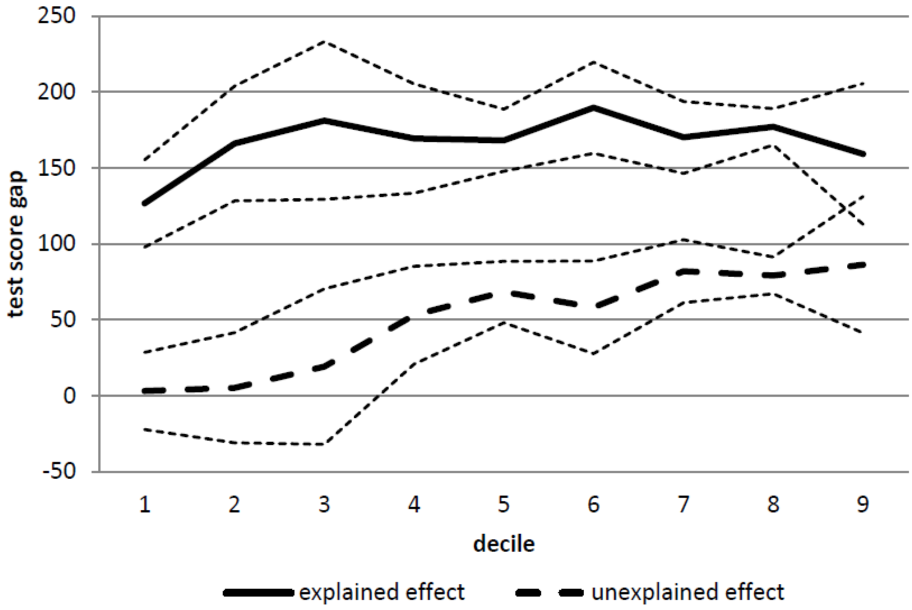
4.4 Decomposition over the test score distribution

Using the estimated counterfactual distribution we are able to determine the relative importance of efficiency and endowment differences for explaining performance differentials over the test score distribution. The respective magnitudes of the explained and unexplained gaps over deciles of the test score distribution with 95 percent confidence bands are shown in figure 2.[25] The explained component increases slightly over the first three deciles after-which it remains fairly constant. The confidence band around the explained effect indicates that the size of this gap component does not change significantly over the test score distribution. Conversely, the unexplained effect increases steadily over the test score distribution from approximately zero effect over the first third of the distribution to 80 points at the upper third. Confidence intervals further indicate that the unexplained effect becomes significantly larger as we move up the test score distribution. The results suggest that an increasing test score gap over the distribution is driven by differences in the returns structure across the two school groups. This is affirmed by figure A3 of the appendix. Whereas the performance gap at the 1st decile can be fully accounted for by differences in the composition of characteristics, 28 and 32 percent of the gap at the 5th and 9th deciles respectively are accounting for by differences in coefficients. As we move up the test score distribution, students from African language testing schools lag further and further behind students from English/Afrikaans schools given higher efficiency in transforming schooling inputs in the latter.

The counterfactual we have tested thus far has allowed for a change in coefficients holding the covariate distribution of African language schools unchanged. We may be interested to investigate the counterfactual distribution derived through combining the conditional distribution in group 0 with the covariate distribution in group 1; that is, $F^1_{Y_0}(y)$. This allows us to assess how the test score distribution of African language testing schools would be affected if they were to possess the student and school characteristics of English and

[25] The confidence bands are based on bootstrapped point-wise standard errors estimated from 100 repetitions. A detailed description of the estimation of the confidence intervals is provided in Chernozhukov et al (2012).

Figure 2: Oaxaca-Blinder decomposition over test score deciles [a]



[a] 95 percent confidence intervals plotted around the estimated effects.

Afrikaans testing schools, yet continue to transform inputs as before. Table 4 indicates the estimated policy effects of estimating this counterfactual, $F^{1}_{Y_0}(y) - F^{0}_{Y_0}(y)$, at each decile. These are compared to the estimated policy effect $F^{0}_{Y_1}(y)-F^{0}_{Y_0}(y)$ from the counterfactual distribution already shown in figure 2. At first glance, the policy effect of changing the covariate distribution compared to the policy effect of changing the conditional distribution (returns structure) is larger at the bottom of the distribution in the case of the former and larger at the top of the distribution in the case of the latter. Therefore, providing a resource rich schooling environment is related to larger improvements in the performance of weaker performing African language schools than more efficient transformation of an already low resource base. Yet, where resource endowments are higher, which is likely to be the case with better performing African language schools, a more efficient transformation of inputs is associated with a greater improvement in test performance. However, closer inspection of the standard errors on these effects indicates no statistically significant difference between the two effects.

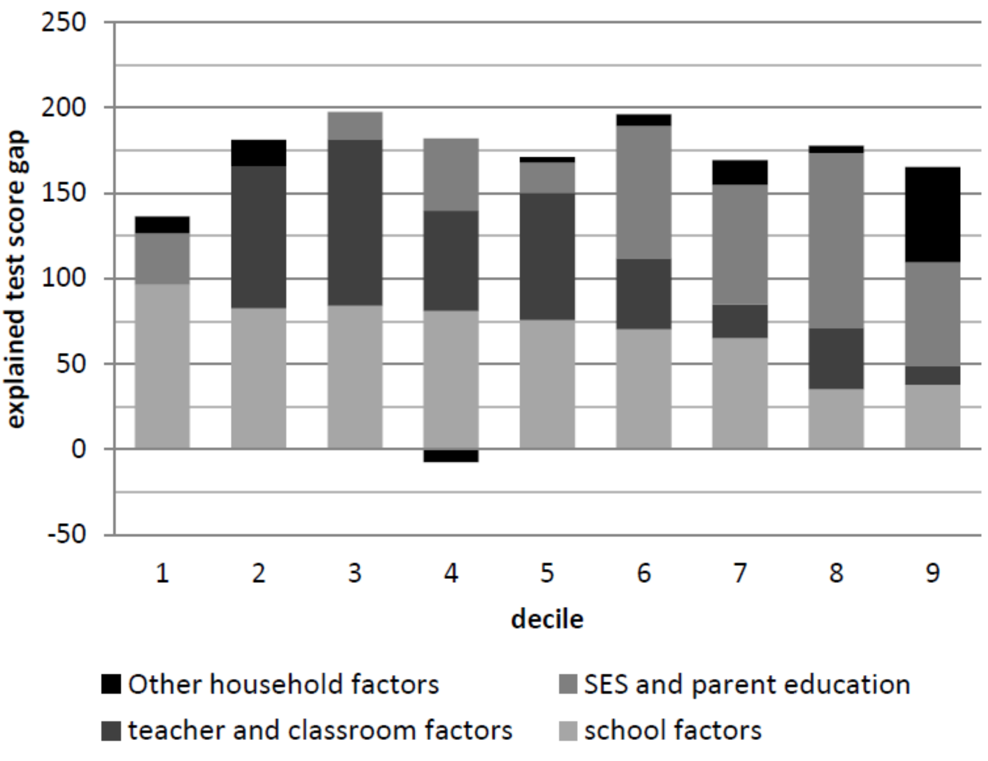
Table 4: comparison of decile policy effects under alternative counterfactuals [a, b]

decile	Policy effect	Point-wise standard error	Policy effect	Point-wise standard error
1	20.093	10.73	3.103	13.00
2	25.907	10.61	5.211	18.51
3	43.029	25.47	19.146	26.09
4	48.912	32.95	53.024	16.43
5	38.129	23.53	68.261	10.30
6	41.102	22.59	58.154	15.55
7	48.789	15.16	81.948	10.55
8	49.058	16.50	79.192	6.16
9	63.401	35.73	86.263	22.89

[a] * p<0.10; ** p<0.05; ***p<0.01
[b] Robust clustered standard errors in parentheses.

Figure 3 presents detailed decomposition results of the explained component across performance deciles. As mentioned, the results are dependent on the sequence in which covariates are included to determine the counterfactual distributions. Here, the sequence is as follows: school factors; classroom and teacher level factors; school SES, household SES and teacher education; and other household factors. Household and school SES are grouped together to represent the affluence of the home and parent background. It is clear that the contributions of the different covariate groups to the explained effect are divergent over the test score distribution. Whilst teacher and school factors explain the larger portion of the performance gap at the bottom end of the distribution, home and socio-economic factors are found to account for a larger proportion of the gap at the upper end of the distribution. Given the detailed decomposition results and the fact that differences in the returns structure has a small, if any, role to play in the performance gap between the weaker performing schools of the two school groups, we would conclude that weaker performing African language schools are at a disadvantaged given poor resources at the school and classroom level. Alternatively, the performance difference between the better performing schools across the two school groups may be accounted for by the dramatically different make-up of the student bodies, which itself may be related to the dissimilar effectiveness of the transformation process of schooling inputs.

Figure 3: detailed decomposition of explained component, across deciles



5. CONCLUDING REMARKS

This study aimed to analyse the performance gap between the students of former Black and former advantaged schools and asks how much of this gap is due to more efficient school processes within the latter group of schools. Using decomposition techniques adopted from the labour literature, the performance gap is decomposed into that part which can be explained by differences in the endowment of educational inputs, and that part which is due to efficiency or quality differences. To this end a nationally representative sample of grade 5 learners was used in which learners were tested in one of 11 official languages. The test language chosen by the school was used to subdivide the data into two groups that proxy the former advantaged and disadvantaged school systems. Specifically, we split schools into those that tested in English or Afrikaans and those that tested in an African language. It is the belief of the author that this division performs better than the typically adopted split based on school wealth as test language appears to proxy closer to the former advantaged (white) school system and shown in figure A2 of the appendix.

An Oaxaca-Blinder decomposition of the average test score gap of 213 points indicated that the larger portion (81.1 percent) of the average gap could be explained by differences in characteristics, with the remaining 18.9 percent due to differences in the coefficients. Detailed decomposition results revealed school SES and parent involvement to be the most important factors contributing to the superior performance of English/Afrikaans schools. It is believed that these variables proxy for a number of organisational, managerial and cultural processes at the school. Oaxaca-Blinder decompositions were further estimated at each decile in the test score distribution using a distribution regression approach. Plots of the unexplained and explained components reveal divergent contributions of characteristics and coefficients to the performance gap over the distribution. The performance gap is estimated to be positive and increasing over the entire distribution, driven primarily by the unexplained component. Better performing students at English/Afrikaans schools are able to extend the gap between them and their peers as a result of higher efficiency gains on their individual and school characteristics.

Although this paper has determined the English/Afrikaans testing schools (as a proxy for the former advantaged school system) to be more efficient than the former Black school system, whether or not a solution can be found to improving the dismal performance of students in former Black schools remains less clear. The results would appear to suggest that bringing the returns to inputs within the Black school system in line with that of the former advantaged school system would serve to improve test performance of learners at the upper end of the distribution. These are likely to be the learners who already attend better endowed African language schools. An alternative policy counterfactual whereby the conditional distribution of African language schools is combined with the English/Afrikaans schools' covariate distribution is also associated with improved performance across the distribution. However, the results suggest that even in a resource rich environment, former disadvantaged schools will not necessarily to be able to effectively transform these inputs into performance.

These counterfactual distributions, of course, assume that the returns structure would be unaffected given a shift in resources. Resultantly, policy advisements based on such counterfactuals can be misleading. Shifting resources from one part of the schooling system to another does not necessarily imply that the two systems will continue to operate in the same way. It is difficult to know whether or not a poor school once endowed with, for example, teachers who adopt performance augmenting classroom practices and an enthusiastic, well-educated parent body will continue to transform these resources as they did when faced with resource constraints. In much the same way, it is difficult to know whether or not, in effect, supplanting the student body of one school system into the schools of another will not necessarily have an effect on the efficiency of those schools.

This paper hoped to add to the debate of whether the poor performance of learners in former Black schools may be accounted for by the poor functioning of these schools or the poor home background of the learners is complex. It seems, at least to the author of this paper, that the efficiency of a school is intimately related to the characteristics of its students, parents, teachers and management. Socio-economic status of a learner and the socio-economic composition of a school play a major determining role in test performance. As suggested, the large and significant coefficients on school SES, parent involvement and teacher qualifications estimated for the group of English/Afrikaans schools may not necessarily reflect the true relationship between these variables and test performance, but partly capture the positive relationship between unobservable factors such as management and accountability and, for example, socio-economic status and test performance. We need to take cognisance of the fact that the results, in particular the increasing unexplained effect over the test score distribution, may be driven by non-random selection of learners and teachers into former advantaged schools. Future analysis would be strengthened by the availability of panel data that would allow researchers to fully control for school quality factors and selection.

REFERENCES

Ammermüller, A. (2007). PISA: What makes the difference? Explaining the gap in test scores between Finland and Germany. *Empirical Economics, 33* (2), 263-287.

Anderson, K., Case, A., & Lam, D. (2001). *Causes and consequences of schooling outcomes In South Africa: evidence from survey data*. PSC Research Report 01-490.

Behrman, J. (1996). Measuring the effectiveness of schooling policies in developing countries: revisiting issues of methodology. *Economics of Education Review*, *15* (4), 345-364.

Blinder, A.S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, *8* (4), 436-455.

Case, A., & Deaton, A. (1999). School inputs and educational outcomes in South Africa. *The Quarterly Journal of Economics*, *3*, 1047-1084.

Case, A., & Yogo, M. (1999). *Does school quality matter? Returns to education and the characteristics of schools in South Africa*. NBER Working Paper No. W7399.

Cattaneo, M.A., & Wolter, S.C. (2012). *Migration policy can boost PISA results – Findings from a natural experiment*. Swiss Coordination Centre for Research in Education Staff Paper 7.

Chernozhukov, V., Fernández-Val, I., & Melly, B. (2012). *Inference on counterfactual distributions*. CeMMAP working papers 05/12, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Chernozhukov, V., Fernandez-Val, I. and Melly, B. (2008). Inference on cdeco Distributions. MIT Department of Economics Working Paper No. 08-16.

Christie, P., Butler, D., & Potterton, M. (2007). Section 2: Literature on "schools that work". In *Ministerial committee: schools that work*, Report to the Minister of Education.

Coleman, J. (1966). *Equality of Educational Opportunity study*. Washington, DC: U.S.

Department of Health, Education, and Welfare.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests, *Psychometrika, 16* (3): 297-334.

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

Da Maia, C.P.P. (2012). *Understanding poverty and inequality in Mozambique: the role of education and labour market status*. (unpublished PhD dissertation). Faculty of Business Management and Economic Sciences, University of Stellenbosch, South Africa.

Desai, Z. (2001). Multilingualism in South Africa with particular reference to the role of African languages in education. *International Journal of Education*, *47* (3-4), 323-339.

DiNardo, J., Fortin, N., & Lemieux, T. (1996). Labour market institutions and the distribution of wages, 1973-1992: a semi-parametric approach. *Econometrica, 64* (5), 1001-1044.

Duncan, K., & Sandy, J. (2007). Explaining the performance gap between public and private School Students. *Eastern Economic Journal*, *33* (2), 177-191.

Filmer, D., & Pritchett, L. (2001). *Estimating wealth effects without expenditure data*. World Bank Working Paper.

Filmer, D., Hasan, A., & Pritchett, L. (2006). *A Millennium Learning Goal: Measuring Real Progress in Education*. Centre for Global Development Working Paper 97.

Fleisch, B. (2008). *Primary education in crisis: why South African schoolchildren under achieve in readin*g. Juta and Company Ltd.

Hanushek, E.A. (1979). Conceptual and empirical issues in the estimation of educational Production Functions. *The Journal of Human Resources*, *14* (3), 351-388.

Hanushek, E.A. (2007). Education production functions. in Steven N. Durlauf and Lawrence E. Blume (eds.), *The New Palgrave Dictionary of Economics*.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, *47* (1), 153-161.

Hernandez-Zavala, M., Patrinos, H., Sakellariou, C., & Shapiro, J. (2006). *Quality of schooling and quality of schools for indigenous students in Gautemala, Mexico and Peru*. World Bank Policy Research Working Paper 3982.

Horrace, W., & Oaxaca, R. (2001). Inter-industry wage differentials and the gender wage gap: and identification problem. *Industrial and Labour Relations Review*, *54*, 611-618.

Jann, B. (2008). *A Stata implementation of the Blinder-Oaxaca decomposition*. ETH Zurich Sociology Working Paper No.5

Jones, F. (1983). On decomposing the wage gap: a critical comment on Blinder's method.

Journal of Human Resources, *18* (1), 126-130.

Juhn, C., Murphy, K.M., & Pierce, B. (1991). Accounting for the Slowdown in Black White Wag Convergence. *Workers and Their Wages*. AEI Press.

Juhn, C., Murphy, K.M., & Pierce, B. (1993). Wage inequality and the rise in returns to skill. *Journal of Political Economy*, *101* (3), 410-442.

Krieg, J., & Storer, M. (2006). How much do students matter? Applying the Oaxaca Decomposition to explain determinants of Adequate Yearly Progress. *Contemporary Economic Policy*, *24* (4), 563-581.

Lam, D., Ardington, C. and Leibbrandt, M. (2011). Schooling as a lottery: racial differences in school advancement in urban South Africa. *Journal of Development Economics, 95* (2), 121-136.

Lemieux, T., (2002). Decomposing changes in Wage Distributions: a Unified approach. *Canadian Journal of Economics*, *35* (4), 646-688.

Machado, J., & Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics, 20*, 445-465.

Neuman, S., & Oaxaca, R.L. (2003). Wage Decompositions with Selectivity Corrected Wage Equations: A Methodological Note. *Journal of Economic Inequality*, *2*, 3-10.

Nielsen, H. (2000). Wage discrimination in Zambia: an extension of the Oaxaca-Blinder Decomposition. *Applied Economics Letters*, *7*, 405-408.

Oaxaca, R. (1973). Male-female wage differentials in urban labour markets. *International Economic Review*, *14* (3), 693-709.

Oaxaca, R., & Ransom, M.R. (1994). On discrimination and the decomposition of wage differentials. *Journal of Econometrics*, *61* (1), 5-21.

Pearson, K. (1901). On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine*, *2* (6), 559–572.

Sakelleriou, C. (2012). *Decomposing the increase in TIMSS scores in Ghana: 2003 – 2007*. World Bank Policy Research Working Paper No. 6084.

Schwab, K. (2012). *The Competitiveness Report 2011-2012*. World Economic Forum, Geneva, Switzerland.

Sohn, K. (2010). The dynamics of the evolution of the black-white test score gap. *Education Economics*. doi:10.1080/09645292.2010.488487

Sohn, K. (2010). A new insight into the gender gap in math. *Bulletin of Economic Research*. doi:10.1111/j.14678586.2010.00358.x

Soudien, C. (2004). Constituting the class: an analysis of the process of integration in South African schools. In Chisolm, L. (ed) 2004. *Changing class: education and social change in post-apartheid South Africa*. HSRC Press, Cape Town. Pg. 89-114.

Spaull, N. (2012). Poverty & privilege: primary school inequality in South Africa. *International Journal of Educational Development*. In press.

Taylor, S. and Yu, D. (2008). *The importance of socio-economic status in determining Educational Achievement in South Africa*. Stellenbosch Economic Working Papers

Timæus, I.M., Simelane, S., & Letsoalo, T. (2012): Poverty, race, and children's progress at school in South Africa. *The Journal of Development Studies*. doi:http://dx.doi.org/10.1080/00220388.2012.693168

Timæus, I.M. and Boler, T. (2007). Father figures: the progress at school of orphans in South Africa. *AIDS*, *21* (7), S83-S89.

Van der Berg, S., & Burger, R. (2003). Education and socio-economic differentials: a study of school performance in the Western Cape. *The South African Journal of Economics*, *71* (3), 496-522.

Van der Berg, S., & Louw, M. (2007). *Lessons learnt from SACMEQII: South African student performance in regional context*. Stellenbosch Economic Working Papers 16/07.

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

Van der Berg, S. (2008). *How effective are poor schools? Poverty and educational outcomes in South Africa*. Center for European, Governance and Economic Development Research (cege) Discussion Papers 69.

Webbink, D. (2005). Causal effects in education. *Journal of Economic Surveys*, *19* (4), 535 560.

Woolman, S., & Fleisch, B. (2006). South Africa's unintended experiment in school choice: how the National Education Policy Act, the South African Schools Act and the Employment of Educators Act create the enabling conditions for quasi-markets in schools. *Education and Law*, *18* (1), 31-75.

Yun, M. (2005). A simple solution to the identification problem in detailed wage decompositions. *Economic Inquiry*, *43* (4), 766-772.

Yun, M. (2008). Identification problem and detailed Oaxaca Decomposition: A general solution and inference. *Journal of Economic and Social Measurement*, *33*, 27-38

Appendix

Note 1

1) Missing data on possession items: missing values on household asset ownership were imputed using average possession within each of the 62 explicit strata (according to province and language). Household SES was subsequently estimated using first principal component analysis (PCA) and then standardised to have a mean of zero and a standard deviation of 1. School SES was calculated as the mean household SES within school and also standardised to have a mean of zero and a standard deviation of 1.
2) Missing data on other learner and household characteristics: "missing/unspecified" was grouped as a separate category and a dummy variable coded "1 = missing/unspecified, 0 = otherwise" was included as a control in the regression model. In most cases, the coefficients on these "missing/unspecified" dummy variables were not found to be significantly different from the reference category. Missing data on categorical variables were therefore grouped with the reference category.
3) Missing values on parent education: imputed using the median parental education of the school.

Figure A1: Kernel density distributions of reading scores (weighted), by school type

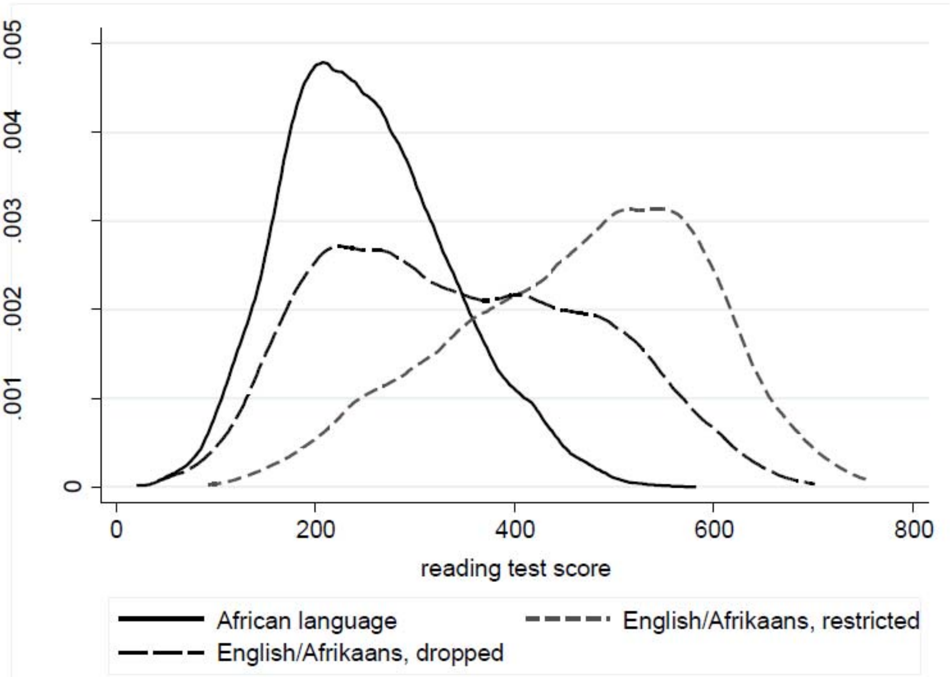


Figure A2: Kernel density distributions of reading scores (weighted), by former department

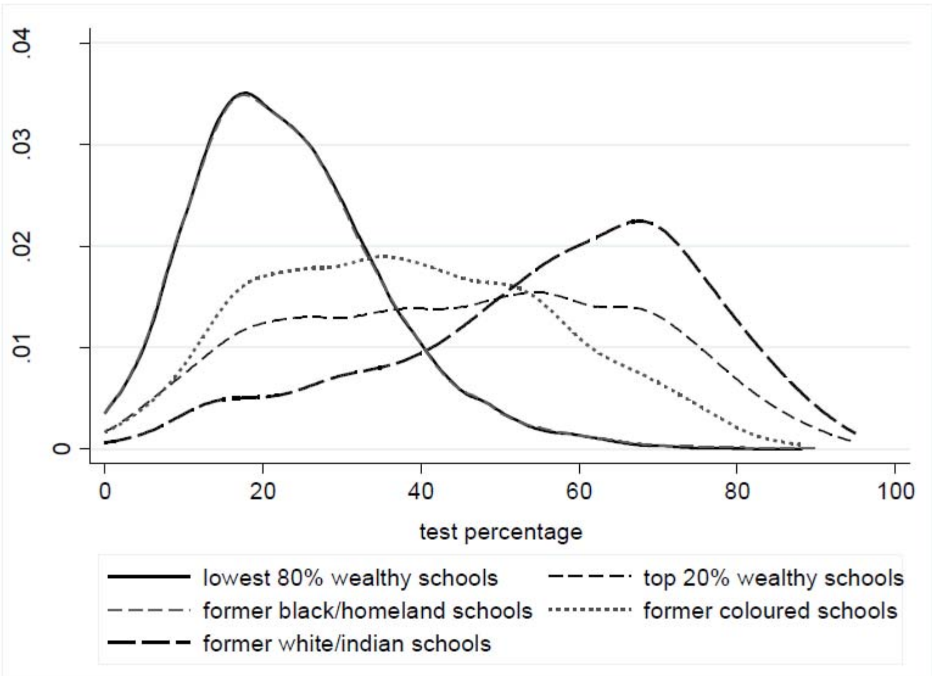


Table A1: Descriptive statistics (weighted) [a]

Variable	Description	English/Afrikaans testing schools		African language testing schools		Difference
		mean	s.e.	mean	s.e.	
Pupil/household:						
Overage for grade 5	Dummy (0,1)	0.165	0.020	0.539	0.018	-0.374***
Underage for grade 5	Dummy (0,1)	0.051	0.008	0.071	0.007	-0.019*
Female	Dummy (0,1)	0.526	0.020	0.513	0.008	0.013
Speaks English regularly at home	Dummy (0,1)	0.557	0.018	0.532	0.021	0.025
Speaks English sometimes at home	Dummy (0,1)	0.332	0.020	0.154	0.008	0.178***
Watches more than 5 hours of television a day	Dummy (0,1)	0.296	0.023	0.348	0.012	-0.052**
Spends more than 5 hours a day playing games on the computer	Dummy (0,1)	0.211	0.013	0.201	0.009	0.009
Parent/s help with reading homework	Dummy (0,1)	0.315	0.023	0.171	0.008	0.144***
Receive reading homework more than once a week	Dummy (0,1)	0.403	0.036	0.346	0.012	0.056
Spends more than an hour of reading homework	Dummy (0,1)	0.179	0.016	0.177	0.008	0.002
Borrows books in home language outside of school	Dummy (0,1)	0.443	0.035	0.238	0.010	0.205***
Mother has at least a matriculation qualification	Dummy (0,1)	0.700	0.047	0.441	0.028	0.259***
Father has at least a matriculation qualification	Dummy (0,1)	0.738	0.043	0.396	0.026	0.342***
Mother speaks the test language at home	Dummy (0,1)	0.619	0.033	0.453	0.015	0.166***
Parent/s read for more than 5 hours per week at home	Dummy (0,1)	0.185	0.011	0.125	0.007	0.060***
High level of early reading activity [b]	Dummy (0,1)	0.523	0.023	0.394	0.009	0.129***
Household socio-economic status index	Continuous (mean = 0, s.d. = 1)	0.778	0.074	-0.269	0.041	1.047***
More than 10 books in the household	Dummy (0,1)	0.651	0.031	0.329	0.017	0.323***

Table A1 continued: Descriptive statistics (weighted) [a]

Learner reads magazines on a daily basis	Dummy (0,1)	0.722	0.019	0.657	0.011	0.065***
Both parents work full-time for pay	Dummy (0,1)	0.298	0.026	0.045	0.005	0.253***
One parent works part-time for pay	Dummy (0,1)	0.296	0.015	0.045	0.005	0.253***
Pupil reports doing worksheets in class more than once a week	Dummy (0,1)	0.794	0.026	0.815	0.010	-0.020
Pupil reports answering questions in class more than once a week	Dummy (0,1)	0.290	0.031	0.503	0.012	-0.214***
School:						
School socio-economic status index	Continuous (mean = 0, s.d. = 1)	1.343	0.128	-0.469	0.070	1.813***
Serious absenteeism problem at the school	Dummy (0,1)	0.169	0.051	0.336	0.039	-0.167***
Moderate absenteeism problem at the school	Dummy (0,1)	0.119	0.049	0.186	0.033	-0.067
School located in an urban area	Dummy (0,1)	0.168	0.056	0.159	0.028	0.010
School located in a sub-urban area	Dummy (0,1)	0.474	0.074	0.117	0.024	0.357***
School offers extended learning time to more than 75 percent of learners	Dummy (0,1)	0.116	0.045	0.090	0.022	0.026
High parent involvement at school [c]	Dummy (0,1)	0.937	0.030	0.561	0.041	0.376***
No free or subsidised lunch programme offered	Dummy (0,1)	0.619	0.070	0.330	0.040	0.289***
Class:						
Class size larger than 30 learners	Dummy (0,1)	0.678	0.070	0.802	0.031	-0.124
Reading series used in classroom teaching	Dummy (0,1)	0.552	0.074	0.582	0.040	-0.030
Long books with chapters used in classroom teaching	Dummy (0,1)	0.129	0.053	0.027	0.015	0.102*
High level of teacher collaboration	Dummy (0,1)	0.380	0.070	0.145	0.027	0.235***
Teacher gives reading homework weekly	Dummy (0,1)	0.667	0.069	0.733	0.035	-0.066
Teacher uses worksheets in classroom teaching weekly	Dummy (0,1)	0.759	0.059	0.869	0.026	-0.110*
Teacher uses group discussion in classroom teaching weekly	Dummy (0,1)	0.486	0.074	0.721	0.036	-0.235***

Table A1 continued: Descriptive statistics (weighted) [a]

Teacher asks learners to give oral feedback of reading weekly	Dummy (0,1)	0.681	0.067	0.812	0.031	-0.131*
Diagnostic testing emphasised in classroom	Dummy (0,1)	0.133	0.050	0.368	0.040	-0.235***
Teacher:						
Teacher has at least a university degree	Dummy (0,1)	0.235	0.063	0.263	0.036	-0.027
Teacher has a post-matriculation diploma	Dummy (0,1)	0.679	0.068	0.497	0.041	0.182**
Male	Dummy (0,1)	0.254	0.059	0.306	0.036	-0.052
Teacher is younger than 30 years	Dummy (0,1)	0.048	0.028	0.014	0.009	0.034
Teacher is 30 to 39 years old	Dummy (0,1)	0.361	0.072	0.454	0.041	-0.092
Teacher is 40 to 49 years old	Dummy (0,1)	0.282	0.067	0.304	0.037	-0.022
Teacher is 50 to 59 years old	Dummy (0,1)	0.290	0.068	0.163	0.031	0.127*
Teacher has less than 6 years of teaching experience	Dummy (0,1)	0.070	0.038	0.099	0.024	-0.029
Teacher has 6 to 15 years of teaching experience	Dummy (0,1)	0.355	0.069	0.503	0.041	-0.148*
Province:						
Western Cape	Dummy (0,1)	0.358	0.070	0.016	0.009	0.343***
Northern Cape	Dummy (0,1)	0.082	0.022	0.002	0.002	0.080***
Free State	Dummy (0,1)	0.023	0.017	0.057	0.013	-0.034
Kwa-Zulu Natal	Dummy (0,1)	0.168	0.063	0.219	0.037	-0.052
North West	Dummy (0,1)	0.030	0.018	0.082	0.020	-0.052*
Gauteng	Dummy (0,1)	0.229	0.067	0.116	0.026	0.113
Mpumalanga	Dummy (0,1)	0.024	0.014	0.091	0.021	-0.068***
Limpopo	Dummy (0,1)	0.012	0.009	0.165	0.026	-0.153***

[a] * p<0.10; ** p<0.05; ***p<0.01
[b] PIRLS generated variable
[c] Parent involvement is coded as taking a value of 1 if the school has more than two formal parent-teacher conferences per year and parents are actively involved in school.

Table A2: WLS estimation results [a, b, c]

Variable	English/Afrikaans testing schools		African language testing schools		Wald test
	coefficient	s.e.	coefficient	s.e.	F-statistic
Overage for grade 5	-36.028***	-5.614	-18.360***	-2.552	9.08***
Underage for grade 5	-8.105	-7.556	-21.969***	-4.119	2.69
Female	22.000***	3.951	26.084***	1.985	0.99
Speaks English regularly at home	23.926***	5.435	12.884***	2.872	3.60*
Speaks English sometimes at home	29.182***	5.506	11.748***	3.398	7.57***
Watches more than 5 hours of television a day	-14.908***	-4.799	-10.344***	-2.177	0.79
Spends more than 5 hours a day playing games on the computer	-12.016**	-3.718	-18.459***	-2.142	2.27
Parent/s help with reading homework	-3.242	-3.981	6.527*	2.698	3.92**
Receive reading homework more than once a week	1.75	4.921	17.784***	2.224	9.42***
Spends more than an hour of reading homework	-11.171*	-4.725	8.245**	2.52	14.39***
Borrows books in home language outside of school	-0.907	-4.081	8.464***	2.25	4.75**
Mother has at least a matriculation qualification	12.255*	5.547	10.640***	3.007	1.46
Father has at least a matriculation qualification	9.931	5.593	9.894***	2.825	0.01
Mother speaks the test language at home	14.913*	6.48	20.285***	3.344	0.56
Parent/s read for more than 5 hours per week at home	5.882	3.575	3.873	3.087	0.43
High level of early reading activity	15.857**	5.009	2.463	2.025	7.67***
Household socio-economic status index	13.996***	3.589	4.326***	1.088	7.31***
More than 10 books in the household	16.476***	4.463	-2.838	2.56	16.67***
Learner reads magazines on a daily basis	3.546	-4.031	5.895**	-2.034	0.60

Both parents work full-time for pay	11.003*	5.077	11.631*	4.596	0.03
One parent works part-time for pay	6.635	4.95	5.571*	2.676	0.05
Pupil reports doing worksheets in class more than once a week	-4.269	4.966	15.939***	3.158	11.43***
Pupil reports answering questions in class more than once a week	-13.052***	3.408	13.045***	2.083	43.34***
School socio-economic status index	55.608***	6.712	80.95*	3.299	36.87***
Serious absenteeism problem at the school	-27.243**	9.093	-5.079	4.639	4.45**
Moderate absenteeism problem at the school	-9.561	9.308	-4.191	6.743	0.13
School located in an urban area	16.502	10.847	24.972*	10.35	0.39
School offers extended learning time to more than 75 percent of learners	36.010***	10.088	7.675	6.762	11.71***
High parent involvement at school	46.919***	13.447	3.638	4.485	8.88***
No free or subsidised lunch programme offered	-4.238	-8.252	5.91	6.302	0.83
Class size larger than 30 learners	-11.42	-8.14	-14.823*	-7.013	0.25
Reading series used in classroom teaching	23.240***	6.631	0.296	3.979	7.74***
Long books with chapters used in classroom teaching	39.746***	7.611	-6.523	-10.379	15.64***
High level of teacher collaboration	13.473*	6.493	9.478	6.77	0.09
Teacher gives reading homework weekly	26.641**	9.637	-4.405	-6.215	7.62***
Teacher uses worksheets in classroom teaching weekly	13.768	7.001	5.882	7.617	0.43
Teacher uses group discussion in classroom teaching weekly	-32.681***	-6.338	16.565**	6.727	23.96
Teacher asks learners to give oral feedback of reading weekly	30.508***	7.265	-3.419	-8.164	8.76
Diagnostic testing emphasised in classroom	-12.342	-7.711	13.888**	4.622	6.73
Teacher has at least a university degree	53.922*	22.268	13.438	8.439	2.66
Teacher has a post-matriculation diploma	34.218	21.931	1.134	7.199	1.87
Male	-6.643	-7.055	8.165	4.64	1.68
Teacher is younger than 30 years	-18.03	-24.628	52.874***	14.313	6.33**

Teacher is 30 to 39 years old	-12.82	-21.847	-8.377	-12.191	0.04
Teacher is 40 to 49 years old	14.684	19.477	-9.71	-9.877	1.49
Teacher is 50 to 59 years old	16.288	21.517	-19.866*	-9.168	2.19
Teacher has less than 6 years of teaching experience	33.101*	15.201	17.79	9.092	0.55
Teacher has 6 to 15 years of teaching experience	14.739*	6.605	4.754	7.898	0.85
Constant	180.473***	36.555	153.612***	16.478	0.48
Observations	2107		9134		
R-squared	0.711		0.333		

[a] dependent variable is the standardised reading score
[b] * p<0.10; ** p<0.05; ***p<0.01
[c] robust clustered standard errors are shown in parentheses. Controls for provincial locations are also included/