Final Report III – Determinants of Grade 12 Pass Rates

as part of the project

Educational Outcomes in South Africa: A Production Function Approach

funded by

SISERA,

The Department of Education

and

The Education Foundation Trust

Haroon Bhorat and Morné Oosthuizen

July 2006

Development Policy Research Unit

School of Economics, University of Cape Town

Private Bag, Rondebosch, 7701

http://www.commerce.uct.ac.za/dpru/

Table of Contents

List of Figures

List of Tables

"There is no blueprint for a model school that can be reproduced and handed out to policymakers, and such a blueprint is unlikely to be developed in the near future" (Hanushek,1995)

1. Introduction

Perhaps the most lasting residue of the system of racial exclusivity in South Africa, lies in the preferential resource allocation and access to schooling provided to White pupils over and above all other racial groups. At the height of apartheid, for every R1.00 spent on White pupils, per capita expenditure on Indian pupils was 76 cents, for Coloured pupils it was 48 cents , while expenditure on each African pupils stood at 19 cents[1]. By 1997, 3 years into the post-apartheid era, the resource allocation had shifted dramatically: Hence for every R1.00 spent on African pupils, 90 and 99 cents on Coloured and Indian pupils – and 71 cents per White pupil (derived from van der Berg, 2006). Although this figure does skew the extent of the reallocation, by virtue of the inclusion of personnel expenditure, it is true that increased spending toward poorer schools has taken place in the post-1994 period (Department of Education, 2003). This fiscal reallocation is captured within the parlance of South Africa, as the process of 'redress' in education. One of the key research issues arising out of this dramatic switch then, is the extent to which this racial restructuring in schooling expenditure has resulted in improved schooling outcomes.

The existence of appropriate microdatasets have enabled fairly detailed studies in particular on the role of schooling in predicting earnings and other labour market outcomes in the post-apartheid South African labour market (Chamberlain & van der Berg, 2002; Anderson, Case & Lam, 2001; Hertz, 2003). However, perhaps in large part due to the paucity of school-specific data, analyses of the various factors that shape schooling outcomes are in short supply for South Africa generally, and even more so for post-apartheid South Africa. In this respect, existing analyses are either dated, not based on national data or attempt to glean schooling outcomes from survey data, rather than schooling datasets (see Crouch & Mabogoane, 2001; Burger & van der Berg, 2003; Case & Deaton, 1999 and Crouch & Mabogoane, 1998).

This study, through the availability of various schooling-related data sets, will attempt to add to what will hopefully become a more expansive literature, focusing on pre-labour market human capital concerns. Our study therefore focuses on Grade 12 pass rates across all Grade 12-offering schools, and in so doing, will provide estimates on the determinants of these pass rates in the post-apartheid period. We rely conceptually and empirically on an achievement production function approach. Section 2 below provides an overview of our method and approach in which we also cover issues of data and the constraints with the surveys utilised. Section 3 provides an overview of the statistics in descriptive form, together with the proposed estimation technique. Sections 4 and 5 provide an overview of the econometric approach, followed by the results and discussion from the estimation. Section 6 concludes.

2. Approach and Method

In utilising a production function approach to estimating schooling outcomes, we immediately enter torturous conceptual and empirical territory. The notion of measuring the impact of inputs into children's education on some metric of schooling outcomes, seems eminently reasonable and indeed, given data for many developing and certainly developed countries, highly attractive. However, as reviews of the results from the plethora of achievement production function studies have indicated, many of these estimates are subject to a variety of econometric problems ranging from omitted variable bias and measurement error to sample selection bias and the presence of unobservables (Glewwe, 2002; Glewwe & Kremer, 2005). Furthermore, Hanushek (1995) examined 96 different studies in the developing world, concerned with the impact of relevant inputs on schooling outcomes – and concludes that we can say very little that is clear and unanimous concerning what resources will lead to improved student performance (Hanushek, 1995:232). In one well-known result, he illustrates that for the 30 studies examining the impact of teacher-pupil ratios on student achievement,

[1] Figures sourced from Moulder (1992) with authors' own computations.

this variable was the correct sign and significant in only 8 of these cases. However, alternatively some have argued that Hanushek (1995) overstates the case, in that he represents statistically insignificant coefficients on resources as evidence that these variables do not matter. Hence, in almost all studies examined, Hanushek's own results suggest that while inputs do not always matter strongly, they do on balance make a difference. Notably, this is not the case for only one variable – namely the teacher-pupil ratio (Kremer, 1996; Glewwe & Kremer, 2005). In this vein then, there remains some possibility that contingent on being able to alleviate some of the econometric problems that arise, the production function approach to understanding schooling outcomes, may yield relevant results.

Our conceptual point of departure therefore, is to understand the determinants of schooling outcomes as a function of at least five vectors. These can be represented in the following manner:

$$A = f(\mathbf{S}, \mathbf{T}, \mathbf{C}, \mathbf{P}, \mathbf{H}) \qquad (1)$$

where A represents some metric of school achievement, which can be for example, standardised test scores or grade attainment. S would represent a vector of school characteristics usually captured by indicators of quality of infrastructure, pupil-teacher ratios and so on. T refers to teacher characteristics and is an attempt at measuring the quality of teachers at a school. C and P are child and parent characteristics respectively, where the former may refer, for example, to innate ability while the latter attempts to capture say, parental tastes for education. Finally, H is a vector of household characteristics that acts as a control for the socio-economic characteristics that are correlated with both the child and parent in respect of the households they reside in. Ultimately then, we are attempting to measure A'(.), as given by the partial derivatives on each of the individual variables defined within the vectors. As will be clearer below, this formulation is independent of the quality and range of the data at our disposable. Indeed many of the potential variables within each of the above vectors are unobservable, with innate ability and parental utility from children's education being the most obvious.

2.1 The Data

The data we have is made up of three distinct components – all of which are national data sets provided to us by the South African National Department of Education (DoE). Firstly, there are records of the aggregate mean Grade 12 pass rates in 2000, for each school whose students sat the examinations. A second data set, the Schools Register of Needs (SRN) for 2000, is a national survey which principally contains data on physical infrastructure and services at schools, but critically has useful background data on for example, whether the school charges user fees, previous apartheid-era department status, the number of teachers (privately- and state-paid) and so on. Through matching the schools' unique identifier (called the Education Management Information Systems – EMIS - number) to geographic co-ordinates, the third and final component of the data was constructed from the Census 2001. Using enumerator area (EA) level data, a limited number of household and community characteristics were matched to schools falling within the relevant EA. Given that the EA contained a maximum of 500 individuals, a close correlation between household, pupils, and schools could be ensured. Ultimately then, the data set contains 5612 schools for which we have their mean matric pass rate, a series of physical infrastructure, services, school- and classroom-type characteristics, household and community characteristics and to a very limited extent, indirect pupil and teacher characteristics.

The data though suffer from a number of drawbacks, within the context of attempting to estimate achievement production functions. Firstly, as is evident, our data is at the school rather than individual-level. Hence intra-classroom variation on the basis of pupil, teacher and parent characteristics is not possible, and this constrains the validity of the estimates derived from the available variables. For example then, the poor coverage of teacher characteristics means that the estimates suffer from omitted variable bias, and in so doing are likely to upwardly bias the coefficients on the school quality variables. In the same vein, the lack of direct parental information again may bias the estimates derived, although indirect parental data is present from the Census data on adult education levels and through information on user fees in the SRN data. Secondly, selection bias is present, and cannot be controlled for. This selection bias is primarily present through the existence of non-zero drop-out rates amongst the grade 12 cohort pupils. Individuals who should have written the Grade 12 examinations, but dropped out earlier and are hence not captured in the pass rate data, would therefore result in an overestimation of the Grade 12 pass rate data. Unfortunately, data on drop-rates by school is not available. Furthermore these drop-rates are likely to be non-randomly distributed, affecting previously disadvantaged schools more and some provinces more than others. Relatedly, there is no reliable information on grade repetition rates. A final cautionary note is that the data does not represent a true post-apartheid set of estimates. This is because apartheid expenditure allocations would be applicable for almost all of these students (controlling for grade repetition) in the pre-1994 period when many (in 1993) would have been in grade 5. So they are imbibed, as it were, with apartheid expenditure allocations from their primary schooling years.

The advantages however, of undertaking such a study is that it is the first nationally representative study of its kind for the post-apartheid period, and could therefore potentially be a guide for improving on future estimates of this sort with better quality data. Secondly, it acts as an important benchmark for assessing the possible impact of redress since 1994, on schooling outcomes. Secondly, unlike many other countries in the developing world, we do have here, a large sample that represents the population of Grade 12-offering schools, which (controlling for the above biases) adds to the robustness of the results. In addition, the use of Census data at a very detailed spatial level allows for household and community characteristics to be fully incorporated into the model in a manner that is almost always absent in similar studies for the developing world.

3. Descriptive Statistics[2]

In trying to provide a more nuanced reflection of the differences in schools, the descriptive and indeed the econometric approach, we provide a brief overview of some of the relevant variables according to the previous education department of the school. This reflects the apartheid-era classification system of African, Indian, Coloured and White schools, and is a reflection, as the data below will show, of the continued racial inequities in the schooling system. The category of 'new' schools refers to those initiated after 1994, although they are constitute the smallest share in the sample. Tables 1 and 2 below therefore present the mean characteristics of some of the key variables under consideration in the study. In the first instance it is clear that our outcome variable, the mean Grade 12 pass rate by school, is significantly lower in former African and Coloured & Indian schools, relative to previously White schools. The pass rate in the latter schools, at 95 percent is close to double that of African schools.

A common feature of the literature is the role of classroom sizes in determining achievement, and this has also been a part of a number of the South African studies (Case & Deaton, 1999; Burger & van der Berg, 2003). The data here suggests similar pupil-teacher ratios for African, Coloured & Indian schools – which are in turn significantly higher than those in former White schools. One key new feature of the post-apartheid schooling system has been the adoption of user fees amongst many schools. We define here, a user fee school as one which employs at least one privately-paid teacher. It is clear though that this is an abiding feature of former White schools, where 92 percent of all these schools charge user fees. Indeed, of all the user fee schools, 30 percent are formerly White schools, despite the fact that they constitute 10 percent of the total sample of Grade 12-offering institutions. As a measure of the continuity in secondary schooling, the lowest grade offered in the specific school is interesting. The contrasts across school types suggest that the mean lowest grade for previously African schools of 7.88 as opposed to 6.34 for former White schools – a 1.5 year schooling difference. Continuity in schooling – with the same school being attended from Grade 1 through to Grade 12 - would ostensibly be an important determinant of engendering positive pass rates.

2 A more detailed descriptive overview of the data is available in the compendium paper to this project, entitled *School Performance in 2000 Matric SC Examinations*, by the same authors.

Table 1: Mean School and Teacher Characteristics, by Apartheid Classification of Schools

Variable/Former Department	***African***	***Coloured /Indian***	***White***	***New***	***Total***
Matric Pass Rate	**48.39**	**74.56**	95.14	48.28	55.47
	(*23.111*)	**(*18.158*)**	(*10.525*)	(*26.288*)	(*26.906*)
Number of Matric Candidates	89.09	**108.94**	103.95	54.76	89.59
	(59.731)	**(*53.985*)**	(*64.119*)	(*43.688*)	(*60.239*)
Pupil-Teacher Ratio	**31.04**	**30.59**	23.72	31.82	30.26
	(*16.912*)	**(*14.022*)**	*(8.127)*	(*12.160*)	(*15.780*)
Schools with User Fees	**0.20**	**0.54**	0.92	0.39	0.31
	(*0.400*)	**(*0.499*)**	(*0.271*)	(*0.489*)	(*0.462*)
Ratio of private:public teachers	**0.03**	**0.201**	0.34	0.24	0.091
	(*0.151*)	**(*2.910*)**	(*0.650*)	(*1.732*)	(*0.988*)
Independent	**0.05**	**0.04**	0.08	0.15	0.05
	(*0.458*)	**(*0.187*)**	(*0.265*)	(*0.360*)	(*0.207*)
Lowest Grade Offerred	**7.88**	**6.78**	6.34	7.73	7.6
	(1.889)	***(2.676)***	*(3.129)*	*(2.645)*	*(2.258)*
Non-std. classroom:learner ratio	**0.0018**	0.0005	0.0006	0.0051	0.0017
	(0.006)	*(0.003)*	*(0.003)*	*(0.013)*	*(0.007)*
Specialist classroom:learner ratio	**0.004**	**0.014**	0.028	0.009	0.008
	(0.010)	***(0.009)***	*(0.042)*	*(0.079)*	*(0.027)*
Principal Office	**0.733**	0.987	0.984	0.540	0.749
	(*0.443*)	(*0.112*)	(*0.125*)	(*0.499*)	(*0.433*)
Chalkboards per classroom	1.02	1.26	1.46	1.12	1.09
	(1.586)	*(0.851)*	*(1.437)*	*(0.746)*	*(1.478)*
Desks per learner	**0.54**	**0.67**	0.91	0.52	0.59
	(*0.444*)	**(*0.469*)**	(*0.490*)	(*0.476*)	(*0.467*)
Chairs per learner	**0.58**	**0.65**	0.98	0.58	0.63
	(*0.555*)	**(*0.494*)**	(1.383)	(*0.635*)	(*0.702*)
At least 1 library	**0.34**	**0.86**	0.91	0.17	0.42
	(0.474)	***(0.350)***	*(0.284)*	*(0.372)*	*(0.494)*
Computer for Teaching & Learning	**0.09**	**0.60**	0.89	0.13	0.22
	(0.287)	***(0.490)***	*(0.313)*	*(0.336)*	*(0.412)*
Telecommunications	**0.72**	0.99	1.00	0.69	0.78
	(0.449)	*(0.092)*	*(0.059)*	*(0.465)*	*(0.417)*
Piped Water indoors	**0.18**	0.69	0.50	0.22	0.25
	(0.380)	*(0.461)*	*(0.500)*	*(0.413)*	*(0.434)*
Electricity for Lighting	**0.71**	0.98	0.98	0.61	0.73
	(0.456)	*(0.129)*	*(0.125)*	*(0.488)*	*(0.442)*
Sports Facilities	**0.62**	**0.84**	0.96	0.54	0.65
	(0.485)	***(0.369)***	*(0.198)*	*(0.499)*	*(0.476)*
Criminal Incident in Previous year	**0.50**	**0.66**	0.63	0.40	0.51
	(0.500)	***(0.473)***	*(0.484)*	*(0.492)*	*(0.500)*
Sample Size	*4,019*	*470*	*565*	*341*	*5,610*
Share of Total	*71.64*	*8.38*	*10.07*	*6.08*	*100*

Source: *Department of Education, 2000; Department of Education, 2000a.*

Notes:
1. Standard Deviations in parentheses. Bold indicates significant difference at 1 or 5 percent level of African and Coloured/Indian school characteristics with previously White school characteristic.
2. Sub-sample of classified schools does not sum to total sample, given that 3.8 percent of sample, or 215 schools were 'unspecified' in terms of their former department status.
3. User Fee School Defined as School with at least one privately paid teacher
4. The share of independent schools refers in part to those schools that have 'converted' from being government schools
5. Non-standard classrooms are those not made of bricks & mortar, while specialist classrooms inlcude laboratories, art & music rooms, woodwork centres and so on
6. Piped water question excludes the unspecified, as in many cases this constituted a large proportion of sample.

Physical learning infrastructure is directly measured through the non-standard classroom- and specialist classroom- learner ratios. African schools yield 0.18 non-standard classrooms per 100 learners compared with 0.06 for White schools. In turn there are on average 0.4

specialist classrooms per 100 learners in African schools, while the figure for White schools is over seven times this, at 2.8. Simply put, the mean African school has a relatively high share of classrooms not made of bricks and mortar, and a relative scarcity of specialist classroom such as laboratories, woodwork rooms and so on.

As an indirect measure for administrative efficiency at schools, and perhaps even as a predictor for attracting quality school heads, the share of schools with principal offices is telling. While about 73 percent of African schools report having a principal office, this figure is close to 100 percent for both Coloured & Indian and White schools. The statistics on in-classroom resources suggests that the discrepancy lies with desks and seats for learners, rather than chalkboards. Hence, we find that while at the mean, two learners are sharing a desk and a chair in former African schools, in former White schools learners invariably have their own chairs and desks. African schools also are unlikely to have a resident library, as only 34 percent report its presence, compared with 86 percent for Coloured and/or Indian schools and 91 percent for former White schools. The differences in physical attributes of the schools for the remaining variables are fairly predictable. There are two results here worth noting. Firstly, that the difference in access to computers for teaching and learning is very high, when compared with other resources categories. Hence while only 9 percent of former African schools had access to at least one computer, the figure for former Whites schools was 89 percent. Secondly, a feature specific to South Africa is the incidence of crime within the school. Data here suggests that while 50 percent of African schools reported at least one crime in the previous year, the proportion was significantly higher for both Coloured/Indian and White schools.

As noted above, through the use of Census 2001, EA-level data we were able to match a number of community characteristics to the area of the specific school. The table below provides a brief descriptive overview of these characteristics, according to the pre-1994 classification code of the school.

Table 2: Mean Household Characteristics by Apartheid Classification of Schools

Variable	*African*	*Coloured /Indian*	*White*	*New*	*Total*
Share Rural	**0.71**	**0.06**	0.08	0.70	0.59
	(0.453)	***(0.234)***	*(0.266)*	*(0.458)*	*(0.492)*
Mean Household Size	**4.75**	4.68	4.37	4.60	4.69
	(3.005)	*(5.345)*	*(7.045)*	*(1.126)*	*(3.777)*
Children per Household	**1.71**	**1.27**	0.82	1.73	1.58
	(0.828)	***(2.042)***	*(0.991)*	*(0.709)*	*(1.041)*
Share of Informal Housing	**34.67**	**4.91**	4.42	40.15	29.60
	(31.892)	***(16.911)***	*(10.164*	*(36.484)*	*(31.717)*
Share of Households with No Piped Water	**29.09**	2.41	1.61	36.22	24.37
	(37.761)	*(10.774)*	*(7.011)*	*(40.883)*	*(36.144)*
Share of Households with No Electricity	**41.89**	**8.42**	5.82	49.50	35.72
	(39.064)	***(18.160)***	*(14.740)*	*(42.203)*	*(38.689)*
Share of Households with No Telephone	**87.19**	**45.51**	37.82	86.93	78.37
	(17.242)	***(22.959)***	*(22.970)*	*(20.118)*	*(42.752)*
Mean Years of Schooling for Adults	**6.62**	**9.26**	10.85	6.45	7.28
	(2.368)	***(1.733)***	*(1.914)*	*(2.610)*	*(2.699)*
Sample Size	*4 019*	*470*	*565*	*341*	*5 610*
Share of Total	*71.64*	*8.38*	*10.07*	*6.08*	*100*

Source: *Statistics South Africa, 2001.*

Notes:
1. Standard Deviations in parentheses. Bold indicates significant difference at 1 or 5 percent level with previously White school area characteristic.
2. Household characteristics are based on enumeration areas from the Census 2001. In cases where the population was under 500 individuals, enumerator areas were combined, on condition that they were part of the same sub-place.

The household data above suggests firstly, that a dominant share of African schools (71 percent) are in rural areas, compared with less than 1 percent of White and Coloured/Indian schools. Both the household size and child dependency ratios are larger, at the mean, in African schools. The four asset deprivation indicators provided – namely access to water, electricity, a telephone and type of housing – all point to significantly greater asset and service vulnerability in areas where former African schools are present. Finally, the one parent-related metric we have –years of schooling of adults – illustrates that adults in African school areas have at the mean incomplete primary schooling, while adults in former White school areas have at least some secondary schooling.

The above descriptive evidence suggests that on a range of school, teacher, parent and community characteristics, there remained in 2000-2001 significant differences between former African schools on the one hand, and former White schools on the other. Of course, the key issue is how these different variables simultaneously impact on the mean Grade 12 pass rates of schools. It is in trying to answer this specific question, that the following section turns to.

4. THE ECONOMETRIC APPROACH

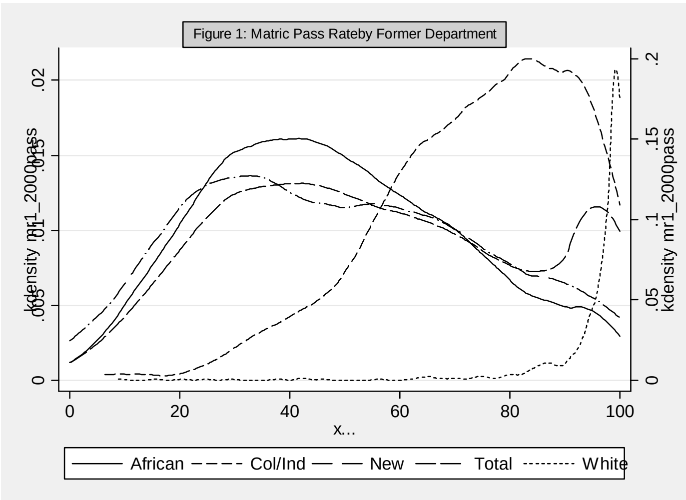
The dependent variable in our model is the mean Grade 12 pass rate by school. The independent variables, as the above has alluded to already, are a range of school and household characteristics drawn from different data sets. Specifically then, the equation we wish to estimate, takes the following generic form:

$$M_i = \alpha + \beta X_i + u_i \qquad (2)$$

where the M_i refers to the mean Grade 12 pass rate in school *i* being a function of the *kx1* vector, *X* of relevant variables, while *β* is the *1xk* vector of parameters. The disturbance term and the constant are captured by u_i and *α* respectively.

A closer inspection of the dependent variable reveals a number of interesting features regarding its distribution. Figure 1 below, provides an indication of this by comparing kernel density estimates of Grade 12 pass rates by former department of the school. There is severe clustering at the top-end of the distribution for former White schools. Specifically, 76.3 percent of all former White schools recorded a pass rate of 95 percent or more in 2000. While the former African school sample and the total sample, approximate a normal distribution, there is a some clustering of grade pass rates from about the 90^{th} percentile onwards, evidenced by the fairly thick tail at the top-end of the distribution. For example, in the case of former African schools, while 2.9 percent (or 118 schools) reported a mean pass rate of 10 percent or less, 5.7 percent (411 schools) returned a pass rate of 90 percent or more.

Figure 1: Matric Pass Rate by Former Department



These features are also true for new schools, where there are suggestions of a bi-modal distribution and a fairly thick tail. Coloured and Indian schools indicate a highly skewed distribution, with a very prominent right-skewing in the distribution. The fact that we may not have well-behaved distributions in the outcome (and indeed other) variables suggests that the estimates on the covariates in our model may yield partial information, should we purely rely on the standard least-squares estimator for our linear model. In this vein then, we propose to estimate our achievement production function, through the use of quantile regressions. First proposed in Koenker and Bassett (1978), quantile regressions refer to the generalised case of the least absolute deviations (LAD) estimator. Hence, while through ordinary least squared estimation, we derive a sample mean through minimising the sum of squared residuals, the sample median can be derived through minimising the sum of absolute residuals (Koenker & Hallock,2001; Koenker & Bassett,1978). If we take a general statement of this approach, across all points, or quantiles, in the distribution we have the estimation for the regression quantile as minimising the equation below:

$$\underset{\beta \in \Re^k}{Min}\left[\sum_{i\in\{i:y_i \geq X_{i\beta}\}} \theta\left|Y_i - X_i\beta\right| + \sum_{i\in\{i:y_i < X_{i\beta}\}} \left(1-\theta\left|Y_i - X_i\beta\right|\right)\right] \tag{3}$$

The above then provides the solution for the *θth* quantile, where $0<\theta<1$, allowing for estimation at any given point in the distribution of the outcome variable. In the above then, Y_i is the dependent variable, x_i is the *kx1* vector of independent variables and β is coefficient vector (Koenker & Bassett,1978). One particular case of the quantile regression is the median regression, which is obtained in the above by setting θ=0.5. Alternative values of θ therefore provide us with different quantile estimates. Ultimately then, while the OLS approach estimates the mean effect of the explanatory variable on the dependent variable, the quantile regression approach enables an estimation at any number of different points in the conditional distribution of the dependent variable.

The application of quantile regression techniques to achievement production functions has at least two antecedents in the literature (Eide,E & M.H Showalter,1999 ; Levin,2001), although both these studies focus on the developed world. No studies, as far as we are aware have applied the estimation technique to a developing, or African, economy. The idea of measuring the impact on the dependent variable, in our case the Grade 12 pass rate, at different points in its conditional distribution, is particularly appealing. Ultimately, it suggests that we will be able to provide a sense of how our explanatory variables influence schools that are at different points in the distribution of pass rates for 2000. In addition, it could hopefully provide a more exact guide to policymakers on how effective expenditure patterns are for schools with different pass rates. For example, one may find that the pupil-teacher ratio, while seemingly important at the mean as a determinant of pass rates, may in fact have different impacts across schools with very high or very low pass rates. This kind of evidence could be invaluable in providing a more nuanced assessment and evaluation of spending patterns on schools.

5. Results and Discussion

Given the above approach, we present the initial results from the estimation in Table 3 below. In Table 3, estimates on the determinants of Grade 12 pass rates, according to different quantiles in the distribution of pass rates are provided. In the table below – the baseline estimates as it were – we run four different equations with three being quantile regressions at the median, and one OLS estimate. Equations (2) through (4) are differentiated by the omission of specific variables. Hence, equation (3) excludes the user fee dummy variable, while equation (4) excludes the previous department dummy, while equation (2) includes both.

The option to exclude the user fee variable, which is effectively a measure of the ratio of public to privately paid teachers, in our preferred estimation, equation 3 below, deserves some brief mention. The user fee variable is very likely to be endogenous through two possible avenues. Firstly, the existence of private paid teachers is suggestive of unmeasured parental preferences for quality schooling: Parents with specific tastes for better schools, will be prepared to pay for it, resulting in a correlation between unmeasured parental preferences and the user fee variable, so biasing the coefficients. More importantly though, the user fee reflects the price of schooling. Its estimate in turn, will have a downward bias, if it is correlated (as is very likely) with some aspects of unmeasured school quality. With few options, in terms of instruments, to control for this endogeneity, we have elected to exclude the variable in our preferred estimation.

Table 3: Achievement Production Function Estimates, 2000

Dependent Variable: Matric Pass Rate	*OLS*	*Quantile Regression Estimates (θ =0.5)*		
	[1]	*[2]*	*[3]*	*[4]*
Pupil- Teacher ratio	0.002	-0.002	0.005	-0.013
User Fee	---	7.119*		
Independent	3.011***	1.316	1.826	-4.460**
Lowest grade	-0.264*	-0.183**	-0.284*	-0.470*
Non-std. classroom:learner ratio	-89.044***	-132.164*	-134.539*	-163.49*
Specialist classroom:learner ratio	-14.393	-25.325*	-22.578*	-0.503
Platoon School	-1.442	-1.583	-2.768	-6.145*
Used for ABET	-0.728	-0.424	-0.582	-1.526
Principal's Office	-0.837	-0.223	-0.596	-2.032***
Accom. For Staff	3.462*	1.710**	2.347*	5.343*
Tuckshop	0.840	-0.018	0.843	1.129
Boards per classroom	-0.075	-0.018	-0.079	0.019
Seats per learner	0.346	0.023	-0.050	-0.253
Desks per learner	1.455*	0.944	1.094	2.383*
Overhead p.l	8.747	3.459	3.714	44.667*
Photocopier p.l	124.346	216.158***	223.967	311.398***
Library	2.634*	2.067*	2.330*	3.159*
Computer for teaching	10.392*	10.200*	11.173*	23.084*
Computer for admin	6.680*	6.889*	7.199*	9.200*
Phone	3.028*	3.663*	3.889*	2.844*
Water Indoors	0.269	-0.636	-0.395	-0.441
Electricity	2.591*	3.459*	3.619*	2.377*
Sports facilities	2.477*	3.223*	3.429*	5.312*
Crime Incident	-2.663*	-2.077*	-2.492*	-2.717*
Col./Indian School	13.039*	14.983*	16.208*	---
White School	26.904*	25.753*	29.159*	---
New School	1.908*	2.512**	3.258**	---
Unspecified Sch.	7.599	13.368*	12.763*	---
Rural	1.162	0.979	1.611	1.307
Household Size	0.134	0.104	0.137	0.269
Children per hh.	-1.067***	-0.426	-0.603	-1.850**
Adult mean yrs of schooling	0.813*	0.630*	0.726*	1.028*
Poverty Index	-0.182	-0.834	-0.800	-1.891*
Constant	35.445*	29.861*	29.718	37.072*
Sample Size	5014	5014	5014	5014
Pseudo R^2	0.440	0.307	0.301	0.271

*: Significant at 1 percent Level
**: Significant at 5 percent Level
***: Significant at 10 percent Level

Notes:
1. Province dummies included, but not shown here.
2. The Household Poverty Index was derived through factor components analysis, using the four variables provided in the Census small-area data on access to water, electricity, telecommunications and the nature of dwelling.
3. The referents for the dummy variables are no user fee; not independent; not a platoon school; not used for ABET; no accommodation for staff; no tuckshop; no library; no computers for teaching or administration; no phone, water indoors or electricity; no sports facilities; no incident of crime; former African school and urban household.
4. Pseudo-R^2 applies to quantile regressions, where Pseudo-R^2= $1-\frac{\text{sum of weighted deviations about estimated quantile}}{\text{sum of weighted deviations about raw quantile}}$

In comparing the OLS results with the preferred equation (3) we are effectively comparing the coefficients at the mean as opposed to the median of the conditional distribution of the dependent variable. The first key result is that the pupil-teacher ratio is insignificant in both specifications – and indeed for the remaining equations as well. This result is consistent with most of the international work, as noted above and is also consistent with studies which apply the quantile regression technique to their data (Hanushek, 1995; Glewwe & Kremer,2005; Levin, 2001 and Eide & Showalter, 1999). Interestingly however, it does contradict some of the studies for South Africa (Burger,R & S.van der Berg,2003; Crouch & Mabogoane, 2001; Case & Deaton,1999), although these studies do differ in terms of foci, data coverage and estimation techniques. Ultimately though, given that this paper is based uniquely on the population of Grade 12 schools – the result that classroom size is not a significant determinant of pass rates is a particularly important one. One caveat ,which we elaborate on below, is that the insignificance of classroom size on pass rates, is not true at all points across the conditional distribution.

The difference in mean and median points in the distribution is shown by the independent school dummy, where the variable is significant and positive at the 10 percent for the mean pass rate, but not for the median. The extent to which continuity in the school is present, through a measure of the lowest grade offered by the school, is clearly a significant determinant of higher pass rates. Hence, the results suggest that reducing the lowest grade offered by 1 year (say from Grade 8 to Grade 7) would increase the median pass rate by 0.28 percent. Put differently, a school that reduced its lowest grade offered by 4 years would, conditional on all other variables, be able to increase its median pass rate by just over 1 percent. This effect is marginally lower at the mean.

The type of classroom variables provide for interesting results. The ratio of non-standard classrooms to learners is significant and negative for both equations (1) and (3), with the effect being 1.5 times larger at the median relative to the mean pass rate. Hence, the higher the ratio of non-standard classrooms in a school, the lower the reported mean or median pass rate. In contrast the mean Grade 12 pass rate is not significantly determined by the proportion of specialist classrooms in a school. The odd result for the median is due to one observation[3]: When the specification is re-run without this school, the coefficient is positive but insignificant. Platoon schools, those that offer multiple teaching hours on the basis of large student numbers and inadequate infrastructure, are an insignificant determinant of median and mean, Grade 12 pass rates in our preferred specifications. The fact that less than 4 percent of the sample reported being a platoon school, may explain the insignificance of this result.

We turn now to the cluster of variables from the presence of an office for the principal to the photocopier per learner estimate, which we term ‘learning infrastructure variables’. It is clear that the results for these variables are weak. The only variable that is significant at the mean and median is that of the provision of accommodation by the school for its staff members. Hence, housing provided by the school to teachers, is positively related to the mean and median pass rate and significant at the 1 percent level, with this effect marginally higher at the mean. The variable may reflect on the ability of the school to attract higher quality teachers: Hence the option of staff accommodation with a high probability of it entailing subsidised living costs, may act as a key incentive for attracting higher quality teachers[4.] The presence of this potential subsidy to teachers therefore serves as one proxy for relatively higher quality teachers being employed by the school. The descriptive statistics suggests that about 58 percent of all former White and 18 percent of all former African schools have staff accommodation facilities. It is important to note however, that even when controlling for previous department status, through our classification dummies, in equations (1) and (3) – the variable remains significant. The combination of a very strong result for the above quality of teacher proxy variable and insignificant learning infrastructure variables, suggests that one of

---

3 This one observation yields a specialist classroom per learner ratio of 1.43, clearly an odd if not incorrect entry. This reflects on the sensitivity of the quantile regression approach, ironically, to outliers if they are sufficiently high (see Rousseeuw & Leroy, 1987). The re-estimation, without this observation does not alter the sign and significance of the other coefficients.

4 Although it is not possible to derive hard data on it, the practice of schools covering utility, accommodation and even food costs for teachers is fairly common amongst those with boarding house facilities for pupils.

the key factors shaping mean and median pass rates in 2000, was the quality of teachers rather than the provision of physical resources within classrooms.

Our next three set of variables we term 'knowledge infrastructure' which measures, specifically, the provision of libraries and computers (for teaching and administrative purposes) within schools. All these three dummy variables are significant at the 1 percent level. In addition they yield very similar magnitudes, whether measured at the mean or median pass rate. It is clear that the presence of a library, at least one computer for teaching and one for administrative purposes are significant in shaping Grade 12 pass rates within schools. In the case of libraries for example, the mean pass rate for schools without a library is 47.3 percent, while for those with a library the figure is 66 percent[5]. The last of these variables – that of computers for administrative purposes – points to the role that increasing administrative efficiency plays in shaping success rates within schools. This has been an argument often made, but seldom empirically verified in the South African literature – namely that many low performing schools are very poorly and inefficiently managed through the lack of administrative (rather than learning) physical and human resources (Crouch & Mabogoane, 2001; Fiske & Ladd,2005).

When examining the importance of utilities and communication to schools – in the form of water , electricity and the presence of a phone – the latter two are positive and significant at the 1 percent level in shaping mean and median Grade 12 pass rates. The importance of telecommunication and electricity is evident, as shown in the descriptive overview above, given that despite the fact that over 70 percent of schools do have these services, the variables remain significant in shaping median and mean Grade 12 outcomes.

The presence of sports facilities are strongly significant at the 1 percent level, and for all specifications estimated. Sport provision for learners remains a strong feature within the South African schooling system, and the relatively high presence of these facilities across all departments is not surprising. Indeed, even for the 10th percentile pass rate schools, 53 percent possess some form of sports facility. There are two possible indirect inferences that can be drawn from this result. In the first instance, schools with privately paid teachers are more often drawn into teaching non-core activities and hence the availability of these facilities may be a marker for the presence of higher quality teachers. Secondly, parents who value a well-balanced, higher quality education for their children may view these facilities as indicative of a better quality school.

South Africa's high national crime levels are also manifest at the level of the school. The results therefore indicate that crime is negatively related to the success of the school. Hence, an incident of crime in the previous year at a school is significant at the 1 percent level, for all specifications, in reducing the mean or median Grade 12 pass rate. The disruption caused by such activity within a learning environment is a key, and possibly unique feature of the South African schooling system.

The final set of schooling variables are classification dummies, which control for the school's previous department status. The referent dummy is a former African school. Hence, former Coloured or Indian, White and (at the median) new schools all perform better than former African schools. Embedded in these dummy variables is of course the legacy of a racially based schooling system which retained an expenditure hierarchy strongly in favour of former White schools. Relatedly though, what these classification dummies do possibly reflect are a large number of omitted variables. Higher quality teachers and wealthier, better educated parents are just two of the teacher and parent characteristics for which we have little data. The classification dummies inevitably, within the South African context, absorb these omitted variables – although clearly as a composite indicator for omitted teacher and parent qualities. These omitted variables together, ensure that Grade 12 pass rates at non-African schools are likely to be significantly higher than the majority of African schools.

The household block in the model, in most cases, returns expected results. Given that the data for each school refers to a maximum of 500 individuals living in the area, the match to the population estimates is likely to be high. Interestingly, location is an insignificant determinant of pass rates. Hence, a rural school is no more or less likely to return different

5 The discrepancy amongst former African (Coloured or Indian) schools is 46 percent (66 percent) for 'no library schools' and 53 percent (76 percent) for 'library schools.

results to an urban school, ceteris parabus. Household size is insignificant, while the number of children is negatively related only to the mean Grade 12 pass rate. The higher the dependency ratio the lower the conditional mean pass rate at the school, although it would appear from these early results that the effect is relatively muted, given its insignificance at the median.

Linked to the above, the mean years of schooling of all adults within the local area of the school, is positive and significant at the 1 percent level. The results suggest that a community with a mean Grade 12 level of education, can increase the median pass rate of the local school by about 8.7 percent. In contrast, a community with a mean of seven years of schooling (or completed primary education), will increase the median pass rate by 5.1 percent. This is a very common result in the literature, namely that parents with higher levels of education are more likely to produce better educated children. In turn, even with school-level data, this obliquely suggests that better educated parents would be more likely to influence the learning environment within the school, given their implicit appreciation of the importance of a high quality learning environment. Finally, it has often been argued within the South African context, that the levels of household vulnerability mitigate very heavily against a schooling system trying to produce good quality graduates. The tentative evidence on the asset poverty index created above, suggests that this interpretation is certainly not true when measuring vulnerability with non-income variables. Hence, the data here suggests that the lack of water, electricity, and a telephone, together with living in informal housing is not a significant determinant of median or mean Grade 12 pass rates in the South Africa.

Ultimately then, the above first set of baseline estimates, yield a number of key results that are important in our understanding of the determinants of Grade 12 pass rates in the South African schooling system. The first is the insignificance of pupil-teacher ratios in determining the mean or the median pass rate – an issue we turn to in greater detail below. The physical infrastructure results are somewhat surprising, in that they suggest that almost all classroom resources are insignificant in shaping pass rates. Rather, it appears to be the presence of staff accommodation, together with non-standard classrooms – which are important in predicting matric outcomes. The results are particularly robust though for knowledge infrastructure (the computer and library variables), where all three variables are significant. Environmental factors such as crime and the availability of services such as electricity and telecommunication are also influential. The strong impact of both the legacy of a racially exclusive schooling system and the importance of omitted variables, is evident in the significance of all the classification dummies. The household variables suggest that while location, asset poverty and dependency ratios are insignificant, the years of schooling of adults in the community, is significantly related to mean and median Grade 12 pass rates.

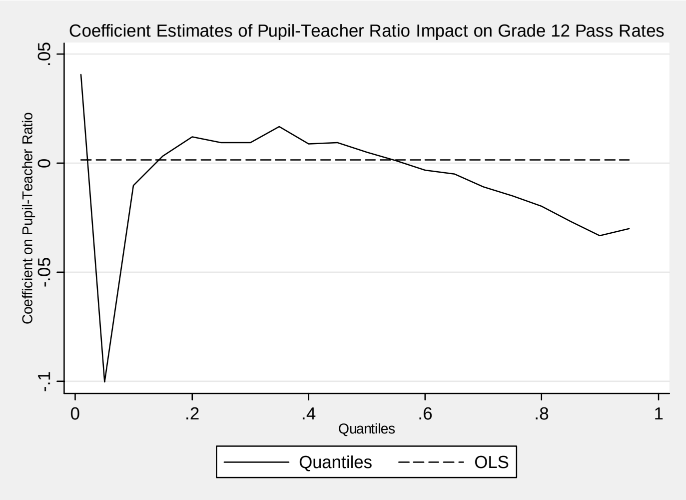
As a caveat to the above though, the magnitude of the significant coefficients also retain a particular importance. Hence, despite the importance of the different variables noted above, there are three groupings of variables that appear to dominate in explaining the impact on Grade 12 performance. Firstly, the coefficient value on the non-standard classroom variable (-134.5 for equation (3)) suggests that the lack of this infrastructure can have a very powerful impact on performance. The classification dummies return the second largest coefficient values for former White and Coloured/Indian schools – reinforcing the role of omitted variables in strongly impacting on pass rates. Finally, both the computer variables report coefficient values are at least double in value when compared with the remaining significant variables. Knowledge infrastructure therefore, is not only a significant, but also a powerful determinant of Grade 12 pass rates.

One of the important questions which arises from the above, is that we are only measuring the impact of the different covariates at the median, while there may be different 'behaviour patterns' in the coefficients when examining different points in the conditional distribution. In trying to deal with this concern, and possibly to add more nuance to the above analysis, we provide below estimates of the determinants of the conditional Grade 12 pass rate, at different percentile (or more generally, quantile) intervals in its distribution. Put differently, Table 4 below estimates the contribution of the different covariates in shaping the 10th, 25th, 75th and

90th Grade 12 pass rates[6]. In this way, we begin to examine why and indeed how, low-, medium- and high-performing schools may differ.

The value of such an exercise is powerfully displayed in our first coefficient on the pupil-teacher ratio. Hence, the ratio is insignificant for the 10th, 25th, 50th and 75th percentile schools, but is significant for the cohort of 90th percentile schools. Closer inspection of the quantiles indicates that the variable is significant from the 80th percentile onwards. The notion therefore that classroom size does not matter needs some qualification and nuance. The variable does not matter, it would appear from this data, for schools in the 0-79th percentiles of the conditional distribution. However for schools from the 80th percentile onwards, classroom size is a significant determinant of pass rates. This variation around the median and mean is illustrated in the graph below, which plots the coefficients derived from Equation (3) above for different points in the conditional pass rate distribution. It is also evident that the magnitude of the ratio increases from the 79th percentile upwards, as the absolute value of the coefficient rises monotonically. Hence, not only is the classroom size variable significant from the upper portion of the distribution onwards, it also increases steadily in its conditional contribution to Grade 12 pass rates. However, it is important to note that the magnitude of the impact of classroom size in this portion of the distribution, remains relatively small. Hence, a reduction in classroom size, for example from 30 to 20 learners in a 90th percentile school, would increase the pass rate by 0.33 percent.

Figure 2: Estimates of Pupil-Teacher Ratio Impact on Grade 12 Pass Rates



Notes:
1. Above are coefficients on Pupil- Teacher ratio estimated according to equation (2) above. In part they are derived from the quantile regressions provided in Equation (3) of Table 3 and those from Table 4 below.
2. The horizontal line refers to OLS estimate as derived from Equation (1) in Table 3 above.

Ultimately then, for the sample we have, and based on the existing data, classroom size is an insignificant determinant of pass rates for all schools up to the 79th percentile of the distribution. From the 80th percentile onwards, classroom size does significantly shape achievement outcomes and increases in magnitude as we move, from this point, up the performance distribution. However, the size of the impact on pass rates remains relatively small.

While being an independent, or private school, was a significant determinant at the mean, it is evident that it is significant only for the sample of 75th percentile schools. The lowest grade offered at the school, significant at the median is also significant and negative at the 10th, 75th and 90th percentile, although lower in value. This suggests that the coefficient produced at the median is not true for only one portion of the conditional pass rate distribution. Both the

[6] Note that the 10th percentile schools does *not* mean that the pass rate for those schools was 10 percent. Instead the relevant pass rates for the 10th, 25th, 75th and 90th percentile schools were 21.43 percent, 33.59 percent, 77.27 percent and 96.67 percent respectively.

classroom type variables provide for interesting contrasts. Hence, the significance of the non-standard classroom variable at the median is true only for the bottom-half of the distribution, where the 10th and 25th percentile schools yield significant and negative estimates. In contrast, the specialist classroom coefficient is insignificant at the 75th and 90th percentile schools, where one would have expected a significant result[7]. Put differently, these results suggest firstly, that while non-standard classrooms do not explain performance at the top-end of the distribution, they are key in explaining poor-performing schools. Secondly, specialist classrooms are unimportant across all points in the Grade 12 pass rate distribution.

The insignificance of the platoon school dummy at the median is also true for the 10th percentile schools, but not so for the 25th percentile, where the coefficient is negative and significant at the 5 percent level. The variable at the 90th percentile however, is oddly positive and significant. Closer inspection of the 90th percentile data suggests that this result may be driven by a few high-performing platoon schools – although the evidence for this is not particularly compelling. As with the median result, none of the coefficients across the percentiles on the ABET school dummy are significant.

In terms of the 'learning infrastructure variables' variables, the results are very mixed, but continue the trend in the above estimates of fairly weak results for most classroom resource variables. Hence in the case of tuckshops, boards per classroom, seats and desks per learner as well as overheads and photocopiers – most quantiles return insignificant results[8]. Within these overall poor results however, there are some interesting issues. Firstly, the principal office dummy is significant but negative for the 75th percentile distribution schools. The descriptive data suggests that high performing former homeland schools could explain this result. Hence, the data illustrates that of the 1 407 schools without a principal's office, 998 (or 71 percent) are former homeland schools. In addition, of these 998 schools, 10 percent are schools which have performed at the 75th percentile or above. Secondly, the significance of the tuckshop variable for the poorest performing schools only, may be an oblique reference to the importance of school nutrition programmes in raising the performance of these schools. Thirdly, the presence of staff accommodation is clearly a robust variable in that it is significant at the 1 percent level at all percentiles in the distribution. This again reinforces our earlier result, namely that this quality-of-teacher proxy variable is critical across all points (except the 10th percentile) of the Grade 12 pass rate distribution. The notion that it is teachers, rather than physical resources, which explain Grade 12 pass rates is strongly supported therefore, by the evidence presented here.

In the case of the 'knowledge infrastructure' variables however, the median results hold true across most percentiles of the distribution. Hence, both computer variables and the presence of a library are positively associated with Grade 12 pass rates across the distribution. Two exceptions are the 10th and 90th percentile coefficients on libraries, which are insignificant.

---

7 The negative and significant coefficient for specialist classrooms at the 25th percentile is due to the one outlier observation, reported earlier. Re-estimation of the specification without this observation returns a coefficient that is positive but insignificant, as with the above estimation. The sign and significance of the other coefficients remained unchanged.

8 When the equation was re-run, without the one outlier variable, the coefficient on the seats per learner variable was positive but insignificant.

Table 4: Achievement Production Function Estimates, by Quantiles

Dependent Variable: Matric Pass Rate	***Quantile (θ) =***			
	0.10	***0.25***	***0.75***	***0.90***
Pupil- Teacher ratio	-0.010	0.009	-0.015	-0.033*
Independent	-1.553	1.641	4.294*	2.705
Lowest grade	-0.166**	-0.025	-0.248*	-0.256**
Non-std. classroom:learner ratio	-219.887*	-107.299**	-19.379	44.605
Specialist classroom:learner ratio	-8.166	-10.782**	-2.251	6.828
Platoon School	-3.091	-4.275**	2.189	3.975**
Used for ABET	-1.008	-0.337	-0.399	-1.330
Principal's Office	0.551	0.800	-2.496*	-1.157
Accom. For Staff	1.890	1.320***	2.455*	2.352*
Tuckshop	3.613*	0.672	-0.603	-1.129
Boards per classroom	-0.011	-0.085	-0.148	0.096
Seats per learner	0.603	-0.172	-0.727***	-0.277
Desks per learner	1.022	1.301***	0.631	0.801
Overhead p.l	6.958	5.399	3.535	22.054*
Photocopier p.l	-337.833*	152.276	529.648*	151.735
Library	1.812	2.568*	1.588**	1.260
Computer for teaching	9.807*	10.389*	9.201*	6.189*
Computer for admin	3.269*	5.128*	9.843*	9.478*
Phone	1.630	2.577*	3.517*	2.795*
Water Indoors	1.007	0.165	-0.351	0.095
Electricity	0.146	3.254*	3.499*	1.594***
Sports facilities	1.395	1.688*	4.095*	2.703*
Crime Incident	-1.967*	-2.059*	-1.525*	-2.847*
Col./Indian School	19.460*	18.568*	11.037*	5.536*
White School	47.229*	41.621*	16.967*	9.660*
New School	1.238	-0.747	3.258*	1.804
Unspecified Sch.	2.895	0.601	11.787*	8.118*
Rural	1.257	0.222	2.739*	1.223
Household Size	0.092	0.028	0.132	0.355*
Children per hh.	-1.037	-0.540	-0.727	-1.529**
Adult mean yrs of schooling	1.009*	0.720*	0.905*	0.579*
Poverty Index	-0.654	-0.512	-0.483	0.495
Constant	13.938*	19.178*	44.794*	67.035*
Sample Size	5014.000	5014.000	5014.000	5014.000
Pseudo R^2	0.237	0.275	0.303	0.220

*: Significant at 1 percent Level
**: Significant at 5 percent Level
***: Significant at 10 percent Level

Notes:
1. Province dummies included, but not shown here.
2. The Household Poverty Index was derived through factor components analysis, using the four variables provided in the Census small-area data on access to water, electricity, telecommunications and the nature of dwelling.
3. The referents for the dummy variables are no user fee; not independent; not a platoon school; not used for ABET; no accommodation for staff; no tuckshop; no library; no computers for teaching or administration; no phone, water indoors or electricity; no sports facilities; no incident of crime; former African school and urban household.
4. Psuedo-R^2= $1 - \frac{\text{sum of weighted deviations about estimated quantile}}{\text{sum of weighted deviations about raw quantile}}$

Schools' access to utilities and communication, again show that access to water is insignificant as a determinant of pass rates – and that this is true across all percentiles in the distribution. The presence of a phone and access to electricity are thus positive and significant across all percentiles, except the 10th, so providing strong confirmation for the median and mean pass rate coefficients.

Another set of variables that are very robust across the pass rate distribution are the presence of sports facilities and the incidence of crime. The variables are correctly signed and significant across almost all percentiles in the distribution, suggesting that they remain key explanatory variables in the achievement production function for Grade 12-offering schools. Likewise the classification dummies are significant in explaining performance at all percentiles of the Grade 12 pass rate distribution, providing strong evidence that omitted variables (principally parental and teacher attributes) explain a significant share of the dependent variable.

The household block results also provide some confirmation for the median and mean estimates. Hence we find no evidence to support the notion that asset or service deprivation within communities is detrimental to school performance. Furthermore, intra-household dynamics, such as household size and child dependency rates appear to matter only for the highest performing schools. It is clear though, that the one community-level factor which is critical is the years of schooling of adults – a proxy for parental education. The coefficient on years of schooling is significant across all reported quantiles. Notably, the impact of an additional year of schooling amongst adults in 90th percentile schools is weaker than in the case of the remaining percentiles. The one additional result that deserves mention is the rural dummy. While it is insignificant across all percentiles, it is significant but positive at the 75th percentile. This result may once again point to importance of rural-based, but relatively high-performing, former homeland schools.

There would seem to be four key conclusions from the above, more nuanced interpretation of the behaviour of the different input coefficients across the conditional distribution. Firstly, in agreement with much of the literature we find that the pupil-teacher ratio is insignificant in explaining the performance of all schools – barring those in the 80th percentile upward. An undue focus on reducing classroom size, at least from the evidence garnered here, would appear to be an unnecessary policy intervention. Indeed, even where the variable is statistically significant its impact on pass rates is fairly small – that is, its substantive significance is low. Secondly, that in conforming with the median and mean results, the evidence points to the relative unimportance of physical classroom resources in explaining school performance. The provision of boards, desks and seats generally seem to matter little in explaining the difference in pass rates amongst Grade 12-offerring schools. There are two caveats to this result: Firstly, specific physical infrastructure for both learning and administration matters in the high performing schools. Hence, we find that the photcopiers per learner ratio has the largest impact on pass rates for 75th percentile schools. In the case of 90th percentile schools, the overheads per learner variable has the highest value coefficient amongst the significant covariates. The second caveat, in terms of physical resources, is that of non-standard classrooms, which again return significant and high value results for the bottom quintile schools. Counter-balancing the tepid physical resource results are the estimates indicating that knowledge infrastructure, together with appropriate access to services and utilities to the school, are critical in explaining relative performance. This is reinforced by the high values of the coefficients for the computer variables, across all points in the distribution estimated above. Thirdly, there is strong, albeit indirect evidence to support the view that while physical resources are relatively marginal, it is teacher and parental characteristics which matter. The significant results for onsite staff accommodation; adult years of schooling and the classification dummies – are all taken as evidence here that the quality of teachers, education level of parents and the latter's tastes for schooling – remain key in understanding the Grade 12 performance of South African schools. The fact, as with the median results, that the classification dummies again retain high values relative to the remaining significant coefficients, are suggestive of the large impact of these omitted variables in explaining pass rates across the conditional distribution.

A final consideration arising out of the above, and indeed overlaying the above two issues – is that of the role played by former homeland schools. These schools would appear to have some role in explaining the odd results obtained for the 75th percentile variables – in particular those for the principal's office dummy and the significance of the rural dummy. While the mean and median pass rates of these homeland schools are very similar to that of former

Department of Education & Training (DET) schools – the evidence does point strongly to a core group of poorly resourced, rural-based high-performing former homeland schools. In particular, it would seem that there are at least 328 former homeland schools, all in rural areas, that are performing above the 75th percentile, constituting some 23 percent of all schools in this sub-sample. The fact that these schools are achieving high pass rates within an environment of low resource commitments, suggests the need for a much closer qualitative and quantitative analysis of these institutions.

5.1 Determinants of Relative Performance

The above results provided an analysis of the impact of different covariates across the conditional distribution of pass rates. It is also useful however to try and estimate the impact of these independent variables on the differences in pass rates. In this vein, we extend the quantile regression approach to estimate a set of inter-quantile regressions, where the dependent variable is the difference between two quantiles. The inter-quantile approach takes the following form:

$$Q_\theta(M_i) - Q_{\theta'}(M_i) = (\alpha_\theta - \alpha_{\theta'}) + (\beta_\theta - \beta_{\theta'})X_i \qquad (4)$$

where Q_θ and $Q_{\theta'}$ refer to the specific quantiles or percentiles for the dependent variable, M_i. The coefficient, $(\beta_\theta - \beta_{\theta'})$ therefore represents the influence of the percentile difference in the covariate on the dispersion in the dependent variable[9]. Given the above, it should be evident that an analysis of selected inter-quantile pass rates is able to provide potentially very useful information on the role our set of independent variables play, in explaining differential pass rates. The table below therefore estimates the impact of the set of covariates on the 90th-10th; 90th–50th and 50th–10th inter-percentile pass rates. Put differently, the coefficients indicate firstly, whether a specific variable is significant or not and secondly, if the variable is significant different from zero, its sign is indicative of whether it increases or reduces the dispersion in pass rates across the two selected percentiles.

If we begin with the household module, what is immediately evident is that all the household variables for all equations estimated, are insignificant. Put differently, this would seem to suggest that socio-economic conditions measured variously by location, years of schooling and dependency ratios are in fact irrelevant in explaining the difference in Grade 12 pass rates. In its most acute form, this result suggests that the dispersion in pass rates between the highest- and lowest-performing schools is not driven in any significant manner by inherited socio-economic factors.

In terms of the pupil-teacher ratio, our nuanced assessment provided above, requires an additional modification. We showed above that the ratio was insignificant in predicting Grade 12 performance below the 80th percentile. However, while the earlier estimates indicated a significant effect at the top of the pass rate distribution, it is evident that the ratio is insignificant in explaining the variability in performance – whether measured by the 90th-10th, 90th-50th or 50th-10th percentile differentials. Put differently, in a policy context, attempts at reducing the most extreme gap - that between 90th and 10th percentile schools through reducing classroom sizes - would have no significant impact on the dispersion of results between the high- and low-performing schools.

In terms of the 90th–10th percentile differences, the independent school dummy which was significant at the mean and 75th percentile is insignificant here. Hence, a school being independent in this particular sample, does not contribute to either increasing or reducing the dispersion in pass rates. This is quite an important result, within the context of the proliferation of independent schools in South Africa, since 1994.

[9] The reported coefficients, as will be shown, are effectively the difference at the respective quantiles across the variables, reported in Table 4. The variance-covariance matrix however, of the inter-quantile regressions, is derived through boostrapping.

Table 5: Inter-Quantile Determinants of Relative Performance

Dependent Variable: Matric Pass Rate	*Inter-Quantile Range*		
	90^{th} - 10^{th}	***90^{th} - 50^{th}***	***50^{th} - 10^{th}***
Pupil- Teacher ratio	-0.023	-0.038	0.015
Independent	4.257	0.878	3.379
Lowest grade	-0.090	0.028	-0.118
Non-std. classroom:learner ratio	264.493**	179.144	85.348
Specialist classroom:learner ratio	14.994	29.406	-14.411
Platoon School	7.066**	6.743**	0.323
Used for ABET	-0.323	-0.748	0.426
Principal's Office	-1.708	-0.561	-1.146
Accom. For Staff	0.462	0.006	0.456
Tuckshop	-4.741*	-1.972**	-2.770**
Boards per learner	0.107	0.175	-0.068
Seats per learner	-0.880	-0.227	-0.653
Desks per learner	-0.220	-0.293	0.072
Overhead p.l	15.096	18.340	-3.243
Photocopier p.l	489.568***	-72.232	561.800*
Library	-0.552	-1.069	0.517
Computer for teaching	-3.618***	-4.984*	1.366
Computer for admin	6.209*	2.280	3.929**
Phone	1.165	-1.094	2.259***
Water Indoors	-0.912	0.490	-1.402
Electricity	1.448	-2.025	3.474
Sports facilities	1.308	-0.727	2.035***
Crime Incident	-0.880	-0.355	-0.525
Col./Indian School	-13.924*	-10.671*	-3.253
White School	-37.569*	-19.499*	-18.071*
New School	0.567	-1.454	2.021
Unspecified Sch.	5.223	-4.644	9.868***
Rural	-0.033	-0.388	0.354
Household Size	0.263	0.218	0.046
Children per hh.	-0.491	-0.926	0.435
Adult mean yrs of schooling	-0.430	-0.147	-0.284
Poverty Index	1.149	1.295	-0.146
Constant	53.097*	37.318*	15.780*
High Quantile R^2	**0.220**	**0.220**	**0.301**
Low Quantile R^2	**0.237**	**0.301**	**0.237**

*: Significant at 1 percent Level
**: Significant at 5 percent Level
***: Significant at 10 percent Level

Notes:
1. Province dummies included, but not shown here.
2. The Household Poverty Index was derived through factor components analysis, using the four variables provided in the Census small-area data on access to water, electricity, telecommunications and the nature of dwelling.
3. The referents for the dummy variables are no user fee; not independent; not a platoon school; not used for ABET;no accommodation for staff; no tuckshop; no library; no computers for teaching or administration; no phone, water indoors or electricity; no sports facilities; no incident of crime; former African school and urban household.

Interestingly, the lowest grade offered, while significant at different points in the pass rate distribution is not significant in explaining the disparity in performance. Non-standard classrooms however remain important in attributing for the gap in pass rates at the extreme ends of the distribution. Specifically the coefficient is positive and significant at the 5 percent level, suggesting that the high share of these non-standard classrooms in the 10^{th} percentile schools serves to widen the dispersion in results when compared with the sample of 90^{th}

percentile schools. This result provides for a powerful policy message – namely that the provision of higher quality physical infrastructure for learning – will significantly reduce the dispersion in Grade 12 performance. Relatedly, the provision of specialist classrooms, would have no significant impact on the difference in relative performance. The presence of a platoon school on the other hand, serves to increase the gap in relative performance between the 90^{th}-10^{th} and 90^{th}-50^{th} percentile schools. It is important to note that this result suggests that, despite the former homeland school result above, the impact of platoon schools at the bottom-end of the distribution is sufficiently strong to increase the dispersion[10].

The presence of a tuckshop yields very strong results, as it is significant and negative across all specifications. Again, this may be some evidence (although this cannot be substantiated with the available data) that school nutrition programmes may be key in reducing the gap in relative Grade 12 performance. As with the previous results however, most classroom resources are insignificant in explaining the difference in performance. Hence, the impact of chalkboards, seats, desks and overhead projectors in explaining the difference in performance across schools is not significantly different from zero.

The stated importance of ensuring administrative efficiency at schools is shown quite strongly in the results here. Hence, the number of photocopiers per learner and the presence of a computer for administrative purposes are both significant and positive in explaining the 90^{th}-10^{th} and 90^{th}-50^{th} percentile differences. The coefficients suggest that a dearth of these facilities at the 10^{th} percentile schools serves to widen the gap in their Grade 12 performance relative to the 90^{th} and 50^{th} percentile schools. Given the import placed on this issue of administrative efficiency and, as noted above, the relative scarcity of evidence on it within schools, this result is particularly noteworthy.

Interestingly, the existence (or lack thereof) of libraries are not significant in explaining the conditional pass rate differentials. On the contrary the presence of a computer for teaching purposes serves to reduce the dispersion in outcomes, by both the 90^{th}-10^{th} and 90^{th}-50^{th} differentials. This would suggest that the provision of computers in 10^{th} and 50^{th} percentile performing schools has served to reduce their variance in performance with the top-achieving cohort. The results for the remaining set of variables ranging from access to services and utilities and the incidence of crime are all fairly weak explanators. Finally though, the classification dummies retain their importance, and again confirming the importance of omitted teacher and parent characteristics in understanding the dispersion in matric performance.

In summary then, the inter-quantile results provide some lessons for attempting to explain the observed differentials in pass rates across schools. Firstly, it is clear that the pupil-teacher ratio remains insignificant even when trying to understand the gap between high- and low-performing schools. Simply put, a 10^{th} percentile school's probability of improved performance does not in any way rest on a reduction in classroom size. Secondly, of all the physical resource variables, it is the provision of standard classrooms and the avoidance of platooning that will reduce the gap in Grade 12 performance – with the former's coefficient though, being over thirty times larger than the latter's. Thirdly, administrative efficiency and knowledge infrastructure matter for reducing the performance gap, and the presence of photocopiers and computers for these purposes is clear evidence for this fact. Importantly, the photcopiers per learner variable has the largest impact on relative performance for both the 90^{th}-10^{th} and 50^{th}-10^{th} pass rate differentials. Finally, unmeasured teacher and parent characteristics, implicitly contained in the classification dummies alludes to possibly the most important set of variables that are likely to explain the gap in Grade 12 pass rates for 2000. Again these set of classification dummies are relatively high in value when compared with the remaining covariates, across the specifications.

[10] Note that if the homeland school effect was stronger, the coefficient would have been negative and significant, suggesting that the presence of platoon schools would reduce the pass rate dispersion.

5.2 Understanding the Dispersion of Results: Actual and Predicted Outcomes[11]

Perhaps a more cogent manner in which to assess the constraints faced by poorer schools is through isolating the impact of inputs of social and resource advantage, as distinct from those variables that are markers of the quality of management, on predicted results. In this vein, we ran a series of regressions, with each estimation isolating measures of social, resource and quality of management advantage. The predicted matric results from this regression were then regressed on the actual results – using the inter-quantile technique concentrating on the 90^{th}-10^{th} percentile differential. The coefficients on these regressions, reported and discussed in detail below, can be thought of as the extent to which resource, social and management quality differentials, assist in reducing the dispersion of results.

Table 6 below then, presents three sets of results emanating from predicted matric results derived from running the 90^{th}-10^{th} percentile inter-quantile regressions on firstly, all the social and resource indices of advantage; on the resource indices and social advantage indicators only and finally on the quality of management variables[12].

Table 6: Predicted and Actual Matric Outcomes: Inter-Quantile Regression Results

Dependent Variable	***Coefficient***	***t-Statistic***
	Social & Resource Advantage	
Predicted Matric Result	-0.412	-10.38
Constant	79.120	42.12
0.9 Pseudo-R^2	0.1850	
0.1 Pseudo-R^2	0.1296	
	Resource Advantage	
Predicted Matric Result	-.3673	-7.44
Constant	77.65	24.67
0.9 Pseudo-R^2	0.1777	
0.1 Pseudo-R^2	0.1183	
	Social Advantage	
Predicted Matric Result	.06021	2.54
Constant	59.76	52.40
0.9 Pseudo-R^2	0.1288	
0.1 Pseudo-R^2	0.0492	
	Management Quality	
Predicted Matric Result	-.7250	-15.31
Constant	95.021	34.89
0.9 Pseudo-R^2	0.1701	
0.1 Pseudo-R^2	0.1702	

Notes: 1. The Social Advantage Index includes the variables rural, household size, children per household, adult mean years of schooling and the asset poverty index.

The data, firstly makes it plain, that all composite measures identified above – be they social, resource or management quality inputs – are significant in affecting the dispersion of results between the 90^{th} and 10^{th} percentile. Importantly, while the social advantage indices individually, appeared to have a fairly benign influence on the dispersion of results, the estimates here indicate that the combination of these measures contributes to a widening of the differential between the best and worst performing schools. Notably however, relative to the forms of advantage (or disadvantage) this indicator does have the smallest impact on relative matric results.

---

[11] Much of this section has drawn on the comments and suggestion from Luis Crouch.

[12] The variables thought to represent a direct or indirect measure of management or the quality of service were: Independent school; principal's office; photocopiers; computers for administration; presence of a phone; previous department status and the province dummies.

Resource advantage, reported a significant coefficient of -0.37, indicating that the supply of physical infrastructure & services to schools can reduce the differential in results. Indeed, the combined impact of resource and social advantage suggests that this provision can help overcome the disadvantage induced by inherited social and economic conditions. This is a key result: namely that policies designed to improve the conditions of schools can in fact overcome inherited social advantage, which is of course much less amenable to policy manipulation. Finally, the quality of management variables also serve to reduce the differential in results - an in fact is the most important of the composite variables. Put differently, it is the quality of the management at schools that has the single largest impact on the dispersion in matric results.

6. Conclusion

The above has attempted, possibly for the first time for South Africa, to estimate an educational production function utilising post-apartheid data that included both schooling and community-level information. In addition, the application of quantile regression techniques allowed for more nuanced information on the determinants of Grade 12 pass rates in 2000. There are a number of key results to emerge from the analysis. These are firstly, that the pupil-teacher ratio is insignificant in explaining pass rates for schools below the 80^{th} percentile. In addition, and critical for policy, the result indicates that the pupil-teacher ratio does not help in explaining the differences in pass rates. Hence, reducing classroom sizes in order to improve the relative performance of schools would be a misplaced intervention.

Secondly, our results show that most physical resources with the exception of the presence of classrooms made of bricks and mortar, are irrelevant in explaining matric performance. This is less true for administration-related resources, which do have a fairly robust explanatory power in predicting performance. Ultimately though, these varied results suggest that the impact of resources on performance is not strong and where there is a significant effect, it is highly dependent on the resource in question, and the metric utilised for the dependent variable. Thirdly, knowledge infrastructure, particularly in the case of computers for teaching, is key in understanding the absolute and relative performance of schools. In addition, the relatively high value of these coefficients, confirms their impact on Grade 12 pass rates.

Fourthly, the strength of the classification dummies and some of the proxy variables for teacher and parent characteristics is a crucial result. For policy purposes, the latter are not easily amenable to intervention, while the former clearly is. The importance of teacher characteristics, admittedly captured rather indirectly here, cannot be doubted and it probably should be the key priority focus for any policy programme aimed at improving Grade 12 performance levels. One redeeming feature, from a policy perspective, of the results is that household vulnerability is a weak predictor of performance. Finally, within the context of future research, it would seem prudent to examine very specific schools, such as for example high-performing relative to low-performing former homeland schools, to get a more detailed analysis of what the key factors and parameters are, which make for high-performing schools.

It is hoped that this initial study can be expanded on through the use of more recent Grade 12 data which also draws on some of the richer data collected on schools. This would ultimately be in a bid to better understand the many and varied predictors of what is clearly a vital component for South Africa's long-run economic success.

7. References

Anderson,K.G, A.Case and D.Lam (2001) Causes and consequences of schooling outcomes in South Africa: Evidence from survey data. *Social Dynamics.* Vol 27(1): 37-59.

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

Case, A and A.Deaton (1999) School inputs and educational outcomes in South Africa. *Quarterly Journal of Economics*. 114(3):1047-1084.

Chamberlain,D & S. van der Berg (2002) *Earning Functions, Labour Market Discrimination and Quality of Education In South Africa.* Stellenbosch Economic Working Paper 02/2002. Department of Economics & Bureau for Economic Research. Stellenbosch University.

Crouch,L and T.Mabogoane (2001) No magic bullets, just tracer bullets: The role of learning resources, social advantage and education management in improving the performance of South African schools. *Social Dynamics.* Vol 27(1):60-78.

Crouch,L and T.Mabogoane (1998) When the Residuals Matter More than the Coefficients: An Educational Perspective. *Studies in Economics and Econometrics.* 22(2)

Department of Education (2000) *Matric Results Data ,2000.* Department of Education: Pretoria.

Department of Education (2001) *Schools Register of Needs Survey,2000.* Department of Education: Pretoria.

Department of Education (2003) *A Review of the Financing, Resourcing and Costs of Education in Public Schools.* Report to the Minister of Education. Department of Education: Pretoria.

Eide,E & M.H Showalter (1999) The effect of School Quality on Student Performance: A Quantile Regression Approach. *Economic Letters*. 58:345-350.

Fiske,E.B & H.F. Ladd (2005) *Elusive Equity. Education Reform in post-Apartheid South Africa.* HSRC Press: Cape Town.

Glewwe,P (2002) Schools and skills in developing countries: Education policies and socioeconomic outcomes, *Journal of Economic Literature,* Vol XL, June: 436-482.

Glewwe,P and M.Kremer (2005) Schools, Teachers and Education Outcomes in Developing Countries. *Handbook on the Economics of Education*. Forthcoming.

Hanushek,E (1995) Interpreting Recent Research on Schooling in Developing Countries. *World Bank Research Observer.* Vol 10(2). August.

Hertz,T (2003) Upward Bias in the Estimated Returns to Education: Evidence from South Africa. *American Economic Review.*93(4). September 2003: 1354-1368.

Kingdon,G (1996) The quality and efficiency of private and public education: A case study of urban India, *Oxford Bulletin of Economics and Statistics,* Vol 58(1):57-82

Koenker,R and G.Bassett (1978) Regression Quantiles. *Econometrica*. Vol 46(1): 33-50.

Koenker,R and K.F.Hallock (2001) Quantile Regression. *Journal of Economic Perspectives.* Vol 15(4). Fall: 143-156.

Kremer,M (1995) Research on Schooling: What We Know and What we Don't – A Comment on Hanushek. *World Bank Research Observer.* Vol 10(2). August.

Levin,J (2001) For Whom the Reductions Count: A Quantile Regression Analysis of Class Size and Peer Effects on Scholastic Achievement. *Empirical Economics.* Vol 26: 221-246.

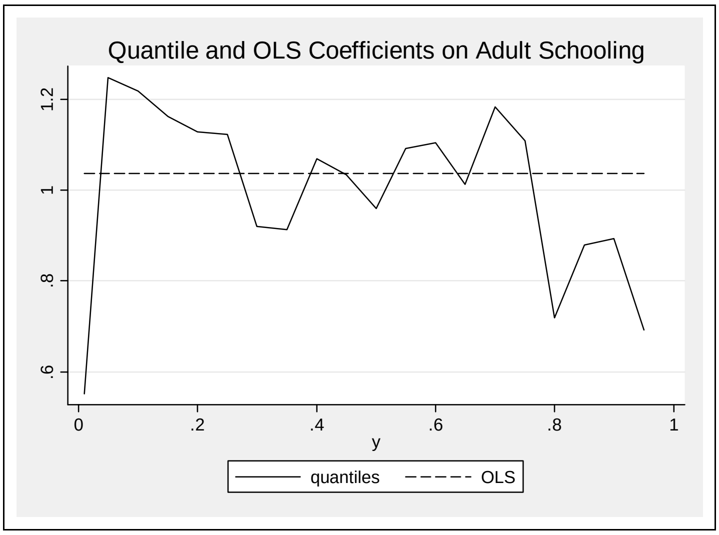
Moulder, J (1992) Education and Distribution In: R.Schrire (ed) *Wealth or Poverty? Critical Choices for South Africa*. Oxford University Press: Cape Town.

Rousseeuw,P.J. and A.M. Leroy (1987) *Robust Regression and Outlier Detection*. New York: Wiley.

Statistics South Africa (2001) *Population Census of South Africa, 2001.* Statistics South Africa: Pretoria.

Van der Berg, S (2006) Public Spending and The Poor Since the Transition to Democracy. In: H. Bhorat and R. Kanbur (2006) Poverty *and Policy in Post-Apartheid South Africa.* HSRC Press: Pretoria.

APPENDIX A



Quantile and OLS Coefficients on Adult Schooling
1.2
1
.8
.6
0
.2
.4
.6
.8
1
y
quantiles
OLS