Determinants of Education Quality

A summary of findings from the Western Cape Primary School Pupil Progress survey of 2003

Commissioned by the Joint Education Trust

Determinants of Education Quality: A summary of findings from the Western Cape Primary School Pupil Progress survey 2003

Introduction

Not much is known about the determinants of education quality in South Africa. Numerous studies have reported that poor and historically disadvantaged communities continue to receive schooling that appears to be inferior in quality. The government has achieved a more equitable distribution of school expenditure and teachers, but the increased resources allocated to schools in previously disadvantaged communities have not induced the expected improvement in education outcomes. We know that despite post-apartheid reforms, predominantly white schools still outperform others. A-aggregates and university endorsements remain concentrated in richer and predominantly white schools, while failures are still considerably higher in predominantly black schools and in poorer schools.

Generally the education data sets available in South Africa do not have sufficient depth to probe the reasons behind the continued poor performance of these previously disadvantaged schools. Traditionally data sets contained variables that captured the impact of socio-economic factors, resources and infrastructure, but included no or little information on vital dimensions such as classroom pedagogy and school management. The rich collection of classroom instruction and school management variables in the Western Cape Primary School Pupil Progress survey 2003 allows researchers to explore the contribution of these dimensions to advancing and constraining improvements in the quality of schooling.

This report gives an overview of this survey, discuss the findings of the survey and describe the potential impact of the findings for education policy in the Western Cape. For the sake of the logical flow of the argument and to keep the document brief, aspects of the analysis that are less important or more technical are only mentioned briefly in the main text and then covered more comprehensively in the appendices.

The data

The Grade 6 survey of primary schools in the Western Cape was undertaken by Joint Education Trust (JET) and the Western Cape Education Department (WCED) towards the end of 2003. The Pupil Progress Project (PPP) survey consists of several modules:

- A **numeracy and literacy test** for pupils in one intact class in each school in the sample.
- The survey also gathered data on the household circumstances and home life of **individual pupils**.
- The **management module** includes detailed school level data on inter alia the curriculum plan and coverage, assessment and learning materials. In each school, field workers interviewed the principal, the head of department and a teacher. For a selection of variables, the survey asked the same question to several of the respondents, allowing a rudimentary consistency check to gauge the reliability of the reported information.
- The management module is complemented by a **classroom observation module**. Teaching practices in both Maths and Language were observed and recorded by trained fieldworkers with a background in education.

The survey was supplemented with two additional data sources:

- Address information was used to link individuals to particular **Census** enumerator areas via Geographic Information System (GIS) to provide additional data on the socio-economic profile of the individual pupil's neighbourhood. Census data from 1996 were employed, because the 2001 Census is not available on enumerator area level due to confidentiality concerns. Although admittedly a crude approximation, the Census averages for the enumerator area gives an indication of the socio-economic position of the pupil's neighbourhood that appears to be meaningful.

- Where deemed useful, variables from the **Western Cape Education Department data set** (e.g. language composition, school fees, former department, poverty of the school community) were added to the survey data set.

An initial sample of 90 schools was selected to be representative of the four Cape Town Metropolitan EMDCs and the Overberg EMDC in the study, within strata defined by three criteria: former department, language of instruction, and three performance categories. Performance was measured relative to expectations. Expectations were predicted using a simple production function and then compared to results from an earlier WCED literacy and numeracy test at Grade 3 level. The sample was devised to be representative of schools, not of pupils in the EMDCs covered in the survey. The same number of pupils was tested in schools of different sizes, thus implicitly assigning an equal weight to schools of different sizes.

Despite prudent research design, there were a number of problems in the final data sets:

- Due to problems experienced in the field, some sampled schools or their replacement schools could not be tested. Three schools refused to participate in the survey and in six other schools testing of pupils could not be concluded in time. Replacements were often not possible due to the encroaching end of year examination period. This reduced the sample of schools with pupil test scores to 81.
- On the individual level, we do not have Census enumerator area information for 1394 of the total 2678 pupils (either because no address was given or otherwise because the given address could not be identified and matched to a Census enumerator area). These observations are discarded in regression analysis, leaving only 1284 of the initial sample of 2678 pupils.
- There are some missing values due to non-response in the school management module. Missing values are considerably higher in the

classroom observation module where variables were imputed from a free-format fieldworker observation of classroom practices.

Analysis shows no evidence of bias resulting from losing nine schools from our sample due to missing test scores, even though the performance of the schools that were dropped were somewhat better than average on the Grade 3 tests earlier conducted by the WCED (see Appendix E.)

For descriptive analysis all available observations were used, resulting in fluctuating samples. The same strategy was followed with multivariate analyses, but here the average sample size was considerably lower because the resulting sample is one for which all variables in the model are available. Sample sizes for the regression analysis mostly ranged between 45 and 60 schools, due to missing values on some variables included in the regressions. This could also introduce bias. The impact of this reduction is considered in Appendix E, which provides strong evidence that non-response and missing values did not introduce significant bias in terms of representivity of schools. The further analysis will consequently be regarded as representative of schools in the five EMDCs covered.

The survey appears to be reasonably reliable, as conclusions from the analysis of this survey are broadly in agreement with findings in the empirical literature. It was possible to gauge the reliability of the reported variables by checking the consistency of those survey questions that were asked to several of the school representatives interviewed. Although there are often contradictions in the answers given by different respondents, the variables appear to be capture something of value. Contrary to initial expectations, many observed variables proved to be more important in our regression analysis as predictors of pupil performance than the reported variables. The significance of the reported variables is encouraging, although this is of course by no means proof of their reliability.

Theoretical approach

The analysis uses a production function approach as point of departure. The model tests whether the performance of pupils is related to the conventional list of educational inputs, including family inputs, peer inputs, school inputs, the pupil's individual characteristics, including innate ability, and proficiency in the language of instruction.

The application of this production function approach to education has been criticized because a production function assumes clearly distinguishable inputs and optimal efficiency in the translation of inputs to outputs. In the education process, it is often difficult to disentangle different inputs. For instance, in the Western Cape poverty is substantially higher among Xhosa-speakers, making it difficult to detect the separate influences of socio-economic circumstances and proficiency in the language of instruction. Despite these shortcomings, the education production function approach has become a standard tool for analysing the effect of different factors on education outcomes.

In our analysis here the focus here falls on identifying the school and classroom level factors. There is little controversy about the important impact of poverty on schooling outcomes, but much uncertainty about the best way to reduce its impact. Based on theoretical considerations, the framework below was constructed to reflect the most prominent potential classroom and school level influences on learning. Due to the cumulative nature of the school quality variable available here (literacy and numeracy test scores for intact classes of grade 6 pupils), the data are expected to be more suited to identifying relationships between school quality and school management than between school quality and the observed classroom practices for their most recent year of schooling.

TABLE 2: Framework of classroom and school level influences

Domain	Level	Construct	Variable	Questions in survey
Instructional	External	Pacing OTL	Coverage	• Management interviews: Is coverage monitored? • Interview with teacher: Are teachers required to submit plans to management? • Interview with teacher: Is coverage

				monitored? • Year plan: Teacher has clear and detailed plan with details on topics, dates and assessment points (cf. teacher not having a plan) • Does the teacher have his/her own copy of the curriculum document(s)?
		Sequence OTL	Coherence	• Does the school distribute textbooks to individual pupils to keep? • Does the school have systems for recovering textbooks?
	Internal	Pacing OTL	Coverage	• Does the teacher use any curriculum document for his or her planning? • Does the teacher use the RNCS curriculum document for planning? • Assessment frequency: How often does assessment occur?
		Sequence pedagogy	Coherence	• Presence of textbooks: Are children given opportunities to engage with textbooks, in or outside the classroom?
		Evaluation	Feedback	• Feedback on assessment: To what extent does the teacher give feedback on assessment tasks?
Regulative		Values	Expectations	• Does the school exhibit a sense of purpose towards teaching and learning during the day? • Does the school have a policy for promoting the cognitive values embodied in the official curriculum (i.e. does it set high expectations for academic achievement)? • Does the principal make reference to setting high academic expectations?
		Time	Time on task	• Does the principal have a master file so that he/she can identify where each teacher should be at any time? • How much time is allocated to maths/language in grade 6?

The empirical literature on school quality in South Africa is weak on casual factors relating to classroom instruction and school management mainly because up to now local data sets have not contained much information about these factors. This report exploits the richness of the 2003 Western Cape Primary School Pupil Progress survey to learn more about the impact of different factors concerning classroom instruction and school management. This richness potentially allows us to move beyond the proxy of former department to investigate the factors that explain the differences in schools performance.

Methodology for estimating model

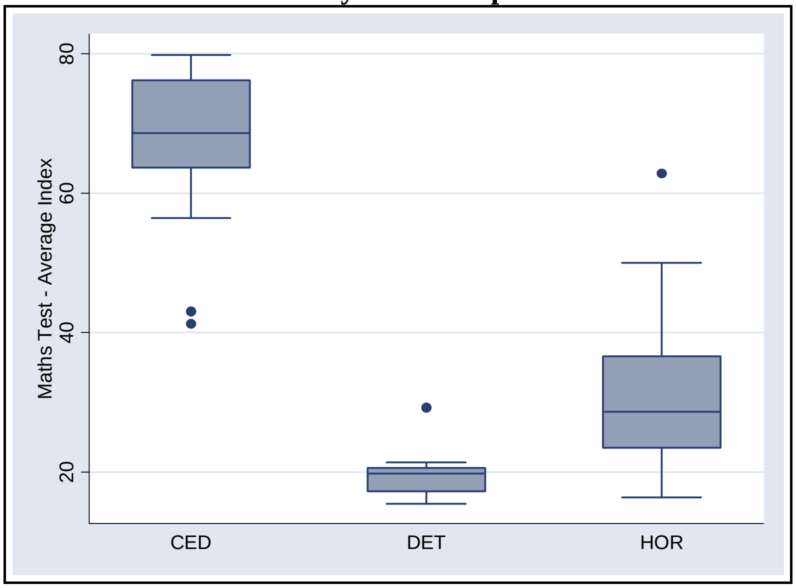
Our empirical analysis investigates determinants of quality in both numeracy and literacy. In each of these cases, we estimate three models: one on the individual level, one on the school level and a last model incorporating the nested structure of the schooling with hierarchical linear modeling. The school level model is most important because it allows us to compare the value of different school inputs, thus providing some guidance for prioritization in education spending and policy making. This is also where most of the variance is located (between schools rather than within schools).

It is difficult to find variables to accurately represent the different influences specified by an education production function. In selection of appropriate variables, we encounter a common problem that information on pupil ability is often not available in surveys. This may bias coefficients of variables that may be correlated with ability.

Another problem is that surveys usually provide information at a point in time, while test scores measure learning that has been accumulated over time. The implicit assumption is then that currently observed school management, classroom practices and family circumstances are correlated to their past levels. School management variables are generally more slow-changing than classroom circumstances and practices and are thus expected to yield stronger results with the cumulative school quality variables.

Furthermore, covariance and the clustering of observations can complicate statistical analysis. Due to the enduring influence of historical factors, there is evidence of such patterns in our data. Schools from the same former department often hzavde similar characteristics also in other respects (e.g. socio-economic status, management or even classroom practices). As an illustration of the clustering, Figure 1 below shows a box-and-whiskers plot of numeracy tests scores by former department. The mean numeracy test scores for ex-CED, ex-HOR and ex-DET schools are 67.0, 30.7 and 19.8 respectively.

FIGURE 1: Box-and-whiskers plots of average mathematics test scores by former department



The detected clustering can complicate statistical analysis by making it difficult to distinguish separate influences on the education process because important school and classroom influences are highly correlated with one another and also with key variables such as poverty, language group and ex-department. Variable values that are highly polarized across clusters may lead to spurious regressions. To deal with these statistical problems associated with clustering, we check the robustness of our models by also estimating it for only ex-HOR schools separately. (There are too few schools in the other two ex-department groups to allow estimation as a separate sample).

The models reported are the product of an iterative elimination process. Due to the extremely high correlation of the school management and classroom variables with former department and the overpowering influence of the latter variable in regression analysis, ex-department is initially excluded from the model to focus on the mechanisms through which advantage and privilege work. The richness of the management and classroom variables in this data set reduces the likelihood of serious omitted variable bias. Once an acceptable model had been identified, it was subjected to repeated testing to ensure stability and the model was also re-estimated limiting the sample to contain only ex-HOD schools. All regressions show robust standards errors.

Findings

Statistical modelling of the factors influencing school performance was done at three levels. For each of numeracy and literacy test results, an analysis was conducted at the school level and at the individual level, and a hierarchical linear model dealing with both the school and individual level was fitted.

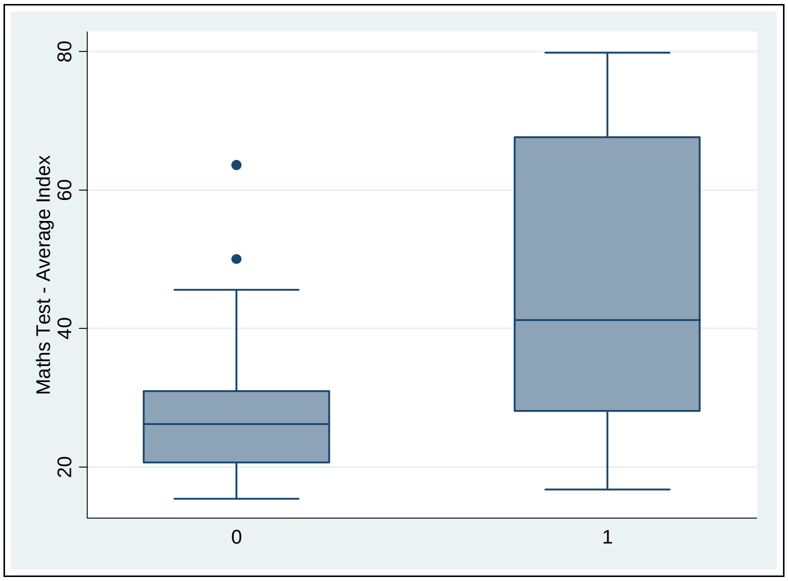
School level numeracy model (Appendix A):

The results for the numeracy test scores confirm findings in the international and local literature regarding the contribution of socio-economic status and teaching resources to school effectiveness. It also allows some insight into classroom and school management variables that are associated with effectiveness. The models predict that, all other things equal:

- Schools that use teacher attendance registers will have an average test numeracy score that is 7 marks higher than schools with no or unutilised attendance registers. The significance of this variable demonstrates the importance of teachers and school discipline.
- Schools that monitor the coverage of submitted teaching plans are likely to have an average numeracy tests score that is 7 marks higher than the test scores of schools that do not monitor teaching plans. This variable may also measure effective school management. The box-and-whiskers plot below shows the dramatic difference in the numeracy test scores for schools that monitor coverage of teaching plans and those who do not (mean of 46 versus 28).
- Where a school had a system for recovering textbooks, this adds an average 7 marks to the test score. This variable is likely to measure both management competence and the availability of textbooks.
- There is a significant impact of a teacher who reports using the RNCS curriculum document for planning, which increased the numeracy test score by 4 marks. This document has more detailed information about sequencing of topics, which is vital for teaching mathematics. Using the RNCS does not

have a significant relationship to measures of management competence and is not significant in the literacy model, providing some support for a narrower and more specific interpretation of this result. Its significance appears to indicate that effective guidance can considerably improve the quality of teaching and learning.

FIGURE 2: Box-and-whiskers plots of average mathematics test scores by monitoring of coverage of submitted teacher plans



In a regression using only the ex-HOR schools as the sample, the coefficients of all variables remain significant and of a comparable size, apart from teacher submission of teaching plans that is no longer significant. This remarkable stability in the results indicate that these results apply not only to schools as a whole, but also specifically to the ex-HOR schools. The R-squareds reported indicate that the model explains a very large proportion of the overall variation, 89%, which stays relatively high (82%) even after omission of the poverty variable.

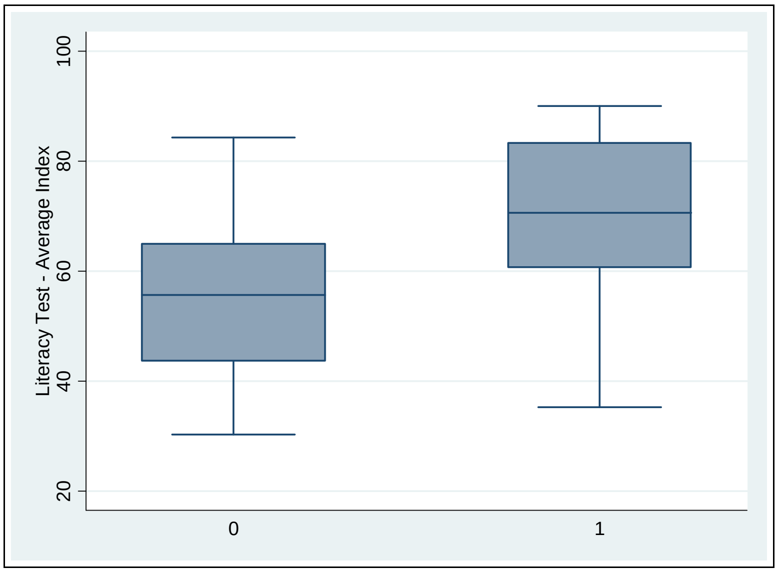
School level literacy model (Appendix A):

The literacy model also shows that poverty constrains learning. The pupil teacher ratio and the average teacher qualification variable (interpreted as representing teacher quality) are not significant, but are retained as standard controls.

The models predict that, all other things equal:

- Systems for recovering textbooks have a large (6 marks) and significant impact on the test scores, if all other factors have been considered. The box-and-whiskers plot in Figure 3 below depicts the large difference in test score for schools with systems to recover textbooks and those without (means 70 vs. 55). As mentioned with the discussion of the numeracy result, this variable may capture both the availability of textbooks and competent school management.
- Monitoring of the coverage of submitted teacher plans is also significant and has an estimated impact of increasing test scores by about 4 marks.
- The set of three dummy variables measuring the availability of the year plan and its level of detail has a sizable and significant influence on test scores. The size of the coefficient increases with the level of detail provided, and a detailed year plan is associated with a gain of as much as 14 marks to the literacy score.
- In addition, the linking of everyday and curriculum knowledge has an impact of increasing the average test literacy test results of a school by 13 marks.

FIGURE 3: Box-and-whiskers plots of average literacy test scores comparing schools with and without systems for recovering textbooks



To investigate the impact of historical factors, ex-DET and ex-HOR indicators were added to the model in turn. Neither of the two ex-department indicators were significant, indicating that the model is capturing a proportion of the mechanisms through which historical factors are working. This moves a step beyond earlier education production function work in South Africa by illuminating the mechanisms that directly impact on the quality of learning rather than only the proximate cause, former department.

The R-squared shows that the regression explains 82% of the overall variance in literacy test results. When the poverty variable is dropped from the sample the explanatory power remains high (74%), possibly partially due the high degree of association between the explanatory variables and poverty.

Comparing school level literacy and numeracy models:

While the models differ starkly at the level of the included variables, there is remarkable agreement between them at the level of the categories. Variables relating to the categories of poverty, school resources, school management, curriculum coverage and textbooks and teaching practices are significant.

Poverty appears to have a stronger influence on numeracy than on literacy test scores. In a simple model with only poverty as predictor, its coefficient is larger in the case of the numeracy model (86) than for the literacy model (67), and a larger part of overall variance (76 versus 63%) is explained for numeracy than for literacy. The poor appear to be more constrained by their school circumstances in terms of becoming numerate than becoming literate.

It is noteworthy that the hours devoted to language and mathematics education are not significantly related to the numeracy and literacy test scores and have not been significant in any of the regressions estimated during the testing process. Its poor performance could be due to misreporting or, alternatively, could be interpreted as evidence that it is not the number of hours invested that matter, but rather how those hours are spent.

Individual level model

The main predictors of individual performance for numeracy and literacy scores are socio-economic circumstances:

- For both literacy and numeracy, the poverty of the area surrounding the school was a significant predictor
- For numeracy, the poverty of the area from which the pupil comes had a small, but detectable additional impact on learning.
- Pupils from a more educated neighbourhood were predicted to perform better. This could also be interpreted as a proxy for the education level of parents.
- Pupils from larger households were expected to perform worse at both numeracy and literacy.
- Pupils from households with a higher ratio of dependents were likely to score lower on the numeracy tests.

There are a few interesting differences between the numeracy and literacy models:

- The literacy model includes more detailed variables on the pupil's understanding of and exposure to the language of instruction.
- Females do significantly better than males (about 5 marks) on literacy, but males perform about 1 mark better on numeracy, although this latter difference is not significant.
- The teacher quality variable is a significant predictor of numeracy, but not literacy test scores.

Other results include that:

- Overage children perform notably worse in both tests.
- Frequent reading and homework make a difference.
- Language is clearly important. Pupils scored considerably higher when the language of the test or the language of instruction was their home language. If it was not their home language, exposure to the language of instruction improved the likelihood of a high score on the literacy test.

The lower proportion of variance explained (R-squared for numeracy model is 0.65 and 0.47 for the literacy model) compared to the school models can be attributed to limited variables available for the individual level model and the importance of unmeasured ability, which explains much individual variation. The only school input variables considered here are teacher quality and teacher-pupil ratios.

Hierarchical Linear Modelling

The breakdown of the variance components shows that the intraclass-correlation rho, the proportion of overall variance that arises from variance in performance between schools, is high at 0.44 for the numeracy and exceedingly high at 0.72 for the literacy test scores. The Kenya SACMEQ II report (SACMEQ 2005: Ch.8, p.14) quotes Willms and Somers (2001) that this value ranged from 19% to 41% for mathematics achievement for Grade 3 and 5 pupils in 13 Latin American countries, whilst Rumberger & Palardy (2003: 14) report a value of 25 % to lie in the normal range. Compared to the magnitudes from three sets of international studies on reading scores covering almost 50 countries, South Africa has by far the highest recorded values. The SACMEQ 2002 rho value of 0.70 for South Africa's reading scores is even exceeded by this dataset. This confirms that inequality in performance between schools in South Africa is exceedingly high.

Hierarchical linear models (HLM) combine the individual and school level models to investigate both between school and within school variation in test results. In our analysis a two-level hierarchical linear model is used. Information about school level is used to predict the slopes and intercept parameters of variables at individual level. For instance, it is possible to allow for the impact of the school's poverty status as well as that of the individual's simultaneously. As starting point for the hierarchical model, the predictors from the individual and school level models are used.

The hierarchical linear models shown in the table in Appendix C are similar to their school and individual level counterparts described in the previous sections.

Although coefficients differ considerably, they are still in the same range. Almost half the variance in test scores between schools can be explained by the HLM models but far less of the individual variance

Both the numeracy and the literacy HLM models include a school level random effect on the intercept and on homework. This indicates that returns to homework are dependent on the school – in fact, extremely so. Table 3 shows that moving from the bottom to the top end of the effort scale (starting to do homework more than three times a week when a pupil previously never did homework) is expected to earn this pupil just more than seven additional marks in the literacy test. It is telling that the model predicts that the same pupil could see a comparable rise in marks (5 marks) if he or she continues to do no homework at all, but moves to a more upmarket neighbourhood and attend a wealthier school. High effort has almost twice the payoff in richer schools (almost 13 marks) compared to poorer schools (7 marks), pointing to the limitations of poorer schools in assisting pupils to escape their socio-economic circumstances.

TABLE 3: Expected literacy score of pupil by effort level, endowment of school and household income

Attending poorer schools	**Low effort**	**High effort**
Pupil from poor household	43.6	50.7
Pupil from rich household	46.8	53.9
Attending richer schools	**Low effort**	**High effort**
Pupil from poor household	45.4	58.0
Pupil from rich household	48.7	61.2

Conclusion

The data set enables researchers to look inside schools and classrooms to identify characteristics and practices associated with effective schools. The regressions indicate that one can often identify effective schools without having to enter the classrooms. Effective schools are characterised by functional teacher monitoring and management systems. The report finds that the availability of textbooks is an essential minimum. Curriculum coverage is vital for ensuring quality of education. Most importantly, the data set allows researchers to statistically start disentangling the causes that lie behind the large differences in performance of pupils in schools from different former departments.

The analysis also shows that individuals are severely constrained by their socio-economic background. To a large extent income and geography remain the most important determinants of the education a young South African will receive. The expected returns from moving from the bottom to the top of the scale in terms of effort invested in homework only marginally outweighs the joint impact of moving to a more affluent neighbourhood and having access to a well-resourced school in this Western Cape sample. Combined with the massive proportion of the total variance in performance that is between schools rather than between individuals in South Africa compared to other developing countries, this makes a very strong case for further work on improving performance in weak schools.

Bibliography

Anderson, Kermyt, Anne Case & David Lam. 2001. Causes and consequences of schooling outcomes in South Africa: Evidence fro survey data. *Social Dynamics*, Vol. 27(1): pp. 37 - 59.

Bernstein, Basil. 1996. *Pedagogy, Symbolic Control and Identity: Theory, Research, Critique.* London: Taylor and Francis.

Bernstein, Basil. 1977. Class, modalities and the process of cultural reproduction: A model. *Language in society*, Vol. 10: pp. 327 – 363.

Case, Anne & Angus Deaton. 1999. School inputs and educational outcomes in South Africa. *Quarterly Journal of Economics*, Vol. 114: pp. 1047 – 1084.

Crouch, Luis & Thabo 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): pp. 60 – 78.

Davidson, Russell & James MacKinnon. 2004. *Econometric Theory and Methods.* Oxford: Oxford University Press.

Hanushek, Eric. 2002. *Publicly provided education.* NBER Working Paper. Cambridge, Mass.: National Bureau of Economic Research.

Rumberger, Russell W. & Gregory J. Palardy. 2003. *Multilevel models for school effectiveness research.* Draft chapter to appear in David Kaplan./ Handbook of Quantitative Methodology for the Social Sciences. Avialable online at: http://www.education.ucsb.edu/rumberger/internet%20pages/Papers/Rumberger%20and%20Palardy--School%20Effectiveness%20chapter%20version%2016.doc

Postlethwaite, T. N. 2004. *Monitoring educational achievement.* Paris: International Institute for Educational Planning.

SACMEQ. 2005. *Kenya SACMEQ II report.* Harare: SACMEQ. Available online at: http://www.sacmeq.org/links.htm

Taylor, Nick, Johan Muller & Penny Vinjevold. 2003. *Getting Schools Working: Research and Systemic School Reform in South Africa*.

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

Appendix A: School Level Model

For the numeracy test scores, we report our model in Table X below. The model includes variables representing the average poverty in the area surrounding the school, school resources, school management and textbooks.

The results confirm findings in the international and local literature regarding the contribution of socio-economic status and teaching resources to school effectiveness. It also allows some insight into classroom and school management variables that are associated with effectiveness. The models predict that, all other things equal, schools that use teacher attendance registers will have an average test numeracy score that is seven marks higher than schools with no or unutilised attendance registers. The significance of this variable demonstrates the importance of teachers and school discipline.

Schools that monitor the coverage of submitted teaching plans are likely to have an average numeracy tests score that is seven marks higher than the test scores of schools that do not monitor teaching plans. This variable may also measure effective school management.

The models also include a variable indicating whether or not the school had a system for recovering textbooks. This adds an average seven marks to the test score. This variable is likely to measure both management competence and the availability of textbooks.

There is also evidence of a significant impact of the teacher reporting that he or she used the RNCS curriculum document for planning. The estimated average impact was an increased four marks on the numeracy test score. The RCNS curriculum document has more detailed information about sequencing of topics, which is vital for teaching mathematics. The RNCS variable has no significant relationship with ex-department or with key management competence measures like section 21 status or availability of a masterfile to identify where a teacher should during the school day. It also has no significance relationship with micro or macro-pacing variables or with the teacher's years of training. Furthermore it is interesting to note that the variable is not significant when included in the model for literacy tests scores.

As argued before, there is evidence that suggests that the influence of historical effects may still be so strong that it is problematic to include these different schools in one regression. To examine this hypothesis, we compare the coefficient of this model (for the full sample) with a regression using only the ex-HOR schools as the sample. The coefficients of all variables remain significant and of a comparable size, apart from teacher submission of teaching plans that now has a p-value of 0.14.

To investigate the impact of historical factors, ex-DET and ex-HOR indicators were added to both numeracy models. The ex-HOR indicator is significant when added to the model, but the ex-DET indicator is not significant.

The R-squareds reported provide an indication of the proportion of the total variation that is explained by these models. The two models explain a very large proportion of the overall variation, 89%, which stays relatively high (82%) even after the omission of the poverty variable (a phenomenon which is partly due to covariance presumably). Due to missing values, the regressions include only 58 schools respectively of the 81 schools for which we have test scores. All coefficients remain significant when the two observations with the highest score and the two observations with the lowest score are eliminated.

TABLE x: School level models for numeracy test scores

Category	Variable	Model with RCNS
Poverty of area surrounding the school	Average poverty of school area	-48.15 (7.60)
School resources	Teacher quality: Average teacher qualifications	3.75 (2.08)
School resources	Pupil-teacher ratio	-0.77 (0.19)
????	Was the teacher attendance register filled in on the day of the visit?	6.75 (3.31)
Pacing OTL	Does the teacher use the RNCS curriculum document for planning?	3.90 (2.00)
Pacing OTL	Interview with teacher: Are teachers required to submit plans to management?	5.99 (2.51)

Pacing OTL	Management interview: Is coverage of submitted teaching plans monitored?	6.55 (1.71)
Sequence OTL	Does the school have systems for recovering textbooks?	5.77 (1.98)
Constant		20.50 (28.62)
R-squared		0.89
Number of observations		58

In agreement with previous research, the literacy model shows that poverty can constrain learning. The pupil teacher ratio and the average teacher qualification variable (interpreted as representing teacher quality) are not significant, but are retained in the model because they are viewed as standard controls in the literature.

As was the case in the numeracy model, systems for recovering textbooks are found to have a large (6 marks) and significant impact on the test scores. However, the interpretation of the variable is complicated because it measures textbook availability on an institutional level, so it may capture general effectiveness and efficiency in addition to textbook availability. The monitoring of the coverage of submitted teacher plans is also significant for both the numeracy and the literacy models. Its estimated impact on test scores is an increase of about 4 marks.

The set of three dummy variables measuring the availability of the year plan and its level of detail has a sizable and significant influence on test scores. As expected, the size of the coefficient increases with the level of detail provided.

In addition, there is also a classroom variable that shows up as significant in the literacy model. The linking of everyday and curriculum knowledge has an impact of increasing the average test literacy test results of a school by thirteen marks. The variable was recoded to become binary so that it now differentiates between classrooms where some attempt was made to link everyday knowledge and curriculum knowledge versus classrooms where no such attempt was made or the attempt used irrelevant or misleading examples.

To investigate the impact of historical factors, ex-DET and ex-HOR indicators were added to the model in turn. Neither of the two ex-department indicators were significant, indicating that the model is capturing a proportion of the mechanisms through which historical factors are working. This moves a step beyond earlier education production function work in South Africa by illuminating the mechanisms that directly impact on the quality of learning rather than only the proximate cause, former department.

The comparison of the full model coefficients with the ex-HOR sample coefficients is problematic because there was no variance in the HOR sample for the binary year plan variable comparing no year plan to the presence of a detailed year plan. None of the schools in the HOR sample had detailed year plans. It is thus impossible to estimate a coefficient for this variable in the restricted sample, but dropping the variable makes the comparison unfeasible.

The R-squared shows that the regression explains 82% of the overall variance in literacy test results. When the poverty variable is dropped from the sample the explanatory power remains high (74%), possibly partially due the high degree of association between the explanatory variables and poverty. Due to missing values in the model's variables, the regression included 52 of the sample of 81 schools. The results appear to be robust to outliers.

TABLE x: School level models for literacy test scores

Category	Variable	Coefficient
Poverty of area surrounding the school	Poverty of community surrounding school	-42.12 (10.37)
School resources	Teacher quality: Average teacher qualifications	-4.54 (2.84)
School resources	Pupil-teacher ratio	-0.25 (0.27)
??	Linking school and everyday knowledge: Does the teacher (at least to a limited extent) appropriately link everyday knowledge and curriculum knowledge?	12.72 (2.67)
Pacing OTL	No year plan available vs. rudimentary year plan available	7.66 (3.21)
Pacing OTL	No year plan available vs. basic year plan available (dates and topics)	8.98 (3.43)
Pacing OTL	No year plan available vs. detailed year plan available	13.90 (3.42)
Pacing OTL	Management interviews: Is coverage monitored?	4.39 (2.40)
Sequence OTL	Does the school have systems for recovering textbooks?	5.90

		(2.23)
Constant		128.74 (39.00)
R-squared		0.82
Number of observations		52

Comparison between the numeracy and literacy models is problematic because it requires an assumption of similarity of the testing instruments, the model selection processes and the samples of the two models. The samples of the two models are rarely the same (as is clear from the number of observations) due to missing values of the variables in the model.

Bearing these caveats in mind, it is worth noting that while the models differ starkly at the level of the included variables, there is remarkable agreement between them at the level of the categories. Variables relating to the categories of poverty, school resources, school management, curriculum coverage and textbooks and teaching practices are significant. It is noteworthy that the hours devoted to language and mathematics education are not significantly related to the numeracy and literacy test scores and have not been significant in any of the regressions estimated during the testing process. This is a reported variable, so its poor performance in the analysis could be due to misreporting. Alternatively, it could be interpreted as evidence that it is not the number of hours invested that matter, but rather how those hours are spent.

Appendix B: Individual level model

The model selection process on the individual level was similar to the school level process. The table below lists the variables available, sorted according to broader categories. Although the model selection process is similar to the school level process, it is less complicated because the data sets has fewer variables available for inclusion and the estimation should also not be as sensitive because of the considerably larger sample.

The table below shows the models for the numeracy and literacy scores of pupils. The main predictors of individual performance are socio-economic circumstances. In the case of literacy, only the poverty of the area surrounding the school matters. For numeracy, the poverty of the area surrounding the school mattered, but the poverty of area from which the pupil comes had a small, but detectable additional impact on learning. Other socio-economic indicators also played a role. Pupils from a more educated neighbourhood were predicted to perform better. This variable could also be interpreted as an approximation of the education level of the individual's parents. The model shows that pupils belonging to larger households were expected to perform worse at both numeracy and literacy and pupils belonging to households with higher ratio of dependents were likely to score lower on the numeracy tests.

There are a few interesting differences between the numeracy and literacy models. For instance, the literacy model includes more detailed variables on the pupil's understanding of and exposure to the language of instruction. The positive coefficient of male gender for numeracy and negative sign on literacy is notable. As was the case with the school level models, the teacher quality variable is a significant predictor of numeracy, but not literacy test scores. Although comparisons of the numeracy and literacy models are more defensible than it was on the school level, it remains difficult to interpret these differences across the models because we do not know much about the comparability of the instruments used for testing.

Overage children perform notably worse in both tests. The models show that frequent reading and homework make a difference. Language is clearly important. Pupils scored considerably higher when the language of the test or the language of instruction was their home language. If the language of the test or instruction was not their home language, exposure to the language of instruction improved their likelihood of a high score on the literacy test.

TABLE X: Individual level model for numeracy and literacy test scores

Category	Variable	Model for numeracy test scores	Model for literacy test scores
Individual inputs	Is pupil male?	1.64 (0.92)	-5.06 (1.12)
	Is pupil overage?	-5.48 (1.05)	-8.00 (1.47)
	Frequency of homework done	2.68 (0.54)	2.06 (0.71)
	Frequency of reading at home	3.13 (0.49)	3.55 (0.58)
Peer inputs	Average poverty of school area	-68.81 (3.69)	-45.92 (4.32)
	Average poverty in area around pupil's home		-0.14 (0.06)
Family inputs	Number of household members	-0.95 (0.22)	-0.69 (0.29)
	Dependency ratio		-1.17 (0.62)
	Percentage of adults matriculated in area around pupil's home	11.03 (2.18)	5.77 (3.45)
Language proficiency	Is the language of the test the home language of the pupil?	2.02 (1.24)	
	Is language of instruction your home language?		8.23 (3.97)
	Frequency of use of language of instruction if not home language		2.65 (1.58)
	How often do you watch TV or listen to the radio in the language of instruction?		1.74 (0.80)
School inputs	Teacher quality: Teacher qualifications	5.03 (1.44)	-2.80 (1.93)
	Pupil-teacher ratio	-0.45 (0.07)	-0.56 (0.08)
Constant		3.84 (0.18)	125.47 (4.29)
R-squared		0.65	0.47
Number of observations		980	944

The numeracy and literacy regressions include 980 and 944 pupils respectively (out of a possible 1394 for which we have both test scores and enumerator area information). The R-squared for the numeracy model is 0.65 and it is 0.47 for the literacy model. The lower proportion of variance explained (cf. school models) can be attributed to the paucity of variables used for the individual level model. The only school input variables considered here are teacher quality and teacher-pupil ratios.

Appendix C: Hierarchical Linear Modelling

As indicated earlier, hierarchical linear models (HLM) combine the individual and school level models to investigate both between school and within school variation in test results. In our analysis a two-level hierarchical linear model is used, with the Level 1 observations at the individual level and the Level II observations at the school level (which here also overlaps with the classroom level). Hierarchical linear models avoid aggregation at school level by modelling variation on the individual level and prevent underestimation of the error terms by taking account of the nested/grouped structure of the individual data. Information about higher levels (classroom and school) is used to predict the slopes and intercept parameters of variables at lower (individual) levels. For instance, it is possible to allow for the impact of the school's poverty status as well as that of the individual's simultaneously. As starting point for the hierarchical model, the predictors from the individual and school level models are used. In all cases where random effects are excluded, it was due to the insignificance of variation between schools in the particular equation.

The hierarchical linear models summarised in Table X below are similar to their counterparts described in the previous section. Although coefficients differ considerably, they are still in the same range.

TABLE X: Hierarchical linear models for numeracy and literacy test scores			
Category	**Variable**	**Model for numeracy test scores**	**Model for literacy test scores**
Fixed effects:			
Individual inputs	Is the pupil male?*	1.54 (0.92)	-4.24* (1.32)
	Is the pupil overage?	-6.09 (0.91)	-9.66 (1.49)
	Frequency of homework done*	3.49 (0.66)	3.38 (0.95)
	Frequency of reading at home	3.26 (0.47)	4.06 (0.56)
Peer and family inputs	Average poverty in area around pupil's home	-0.09 (0.04)	-0.16 (0.05)

Language proficiency	Is the language of the test the home language of the pupil?	5.50 (2.00)	
	Is the language of instruction the home language of the pupil?		3.75 (1.53)
Constant*		16.15 (2.60)	42.86 (3.84)
Variance decomposition:			
Number of observations		49 schools 991 pupils	48 schools 994 pupils
Between school variance as proportion of total variance		44%	72%
Explained variance on school level		44%	47%
Explained variance on individual level		24%	14%

* In the case of the literacy scores, the model included level 2 random effects on the intercept and the coefficients of the homework and male indicators. The effective homework coefficient is calculated by adding the coefficient reported here to the significant slope effect (-4.05) times the difference between the particular school's poverty index and the school poverty index grand mean. The numeracy model included a level 2 random effect on the intercept and the homework coefficient.
Note: The full HLM output for both models is appended to this report.

The breakdown of the variance components shows that the intraclass-correlation rho, the proportion of overall variance that arises from variance in performance between schools, is high at 0.44 for the numeracy and exceedingly high at 0.72 for the literacy test scores. The Kenya SACMEQ II report (SACMEQ 2005: Ch.8, p.14) quotes Willms and Somers (2001) that this value ranged from 19.5% to 41.2% for mathematics achievement for Grade 3 and 5 pupils in 13 Latin American countries, whilst Rumberger & Palardy (2003: 14) report a value of 25 % to be "within the range that Coleman found in his 1996 study and the range found in other recent studies of student achievement using similar models". Table X below shows the range of this magnitude from three sets of international studies, arranged based on the reading scores. South Africa has by far the highest recorded values in the almost 50 countries covered, with Namibia its closest rival in terms of this measure of the degree to which inequality occurs between schools. The SACMEQ 2002 rho value of 0.70 for South Africa's reading scores is even exceeded by this dataset, although the value for numeracy is considerably lower. This confirms that inequality in performance between schools in South Africa is exceedingly high.

TABLE X: Proportion of intra-class correlation rho (variance at school level) from PIRLS and SACMEQ I and II studies and from this study (arranged by rho for reading scores)			
Country / territory	**Study**	**Rho for Reading score**	**Rho for Maths score**
Seychelles	SACMEQ II 2002	0.08	0.08
Iceland	PIRLS 2001	0.084	..
Slovenia	PIRLS 2001	0.087	..
Sweden	PIRLS 2001	0.087	..
Norway	PIRLS 2001	0.096	..
Cyprus	PIRLS 2001	0.105	..
Turkey	PIRLS 2001	0.132	..
Germany	PIRLS 2001	0.141	..
Czech Republic	PIRLS 2001	0.157	..
France	PIRLS 2001	0.161	..
Zanzibar	SACMEQ I 1995	0.17	..
Canada (Ontario, Quebec)	PIRLS 2001	0.174	..
England	PIRLS 2001	0.179	..
Scotland	PIRLS 2001	0.179	..
Netherlands	PIRLS 2001	0.187	..
Italy	PIRLS 2001	0.198	..
Latvia	PIRLS 2001	0.213	..
Lithuania	PIRLS 2001	0.214	..
Greece	PIRLS 2001	0.221	..
Hungary	PIRLS 2001	0.222	..
Malawi	SACMEQ I 1995	0.24	..
Slovak Republic	PIRLS 2001	0.249	..
New Zealand	PIRLS 2001	0.25	..
Mauritius	SACMEQ I 1995	0.25	..
Zanzibar	SACMEQ II 2002	0.25	..
Botswana	SACMEQ II 2002	0.26	0.22
Mauritius	SACMEQ II 2002	0.26	0.25
Zambia	SACMEQ I 1995	0.27	..
Zimbabwe	SACMEQ I 1995	0.27	..
Macedonia	PIRLS 2001	0.271	..
Malawi	SACMEQ II 2002	0.29	0.15
Hong Kong	PIRLS 2001	0.295	..
Mozambique	SACMEQ II 2002	0.30	0.21
Zambia	SACMEQ II 2002	0.32	0.22
SACMEQ Total (across all countries)	SACMEQ I 1995	0.33	..
Kuwait	PIRLS 2001	0.334	..
Tanzania	SACMEQ II 2002	0.34	0.26
Bulgaria	PIRLS 2001	0.345	..
Belize	PIRLS 2001	0.348	..
Romania	PIRLS 2001	0.351	..
Swaziland	SACMEQ II 2002	0.37	0.26
SACMEQ Total (across all countries)	SACMEQ II 2002	0.37	0.32
Iran	PIRLS 2001	0.382	..
Lesotho	SACMEQ II 2002	0.39	0.30
Moldova	PIRLS 2001	0.395	..
Israel	PIRLS 2001	0.415	..
Argentina	PIRLS 2001	0.418	..
Kenya	SACMEQ I 1995	0.42	..
United States	PIRLS 2001	0.424	..
Russian Federation	PIRLS 2001	0.447	..
Kenya	SACMEQ II 2002	0.45	0.38

TABLE X: Proportion of intra-class correlation rho (variance at school level) from PIRLS and SACMEQ I and II studies and from this study (arranged by rho for reading scores)			
Country / territory	**Study**	**Rho for Reading score**	**Rho for Maths score**
Colombia	PIRLS 2001	0.459	..
Morocco	PIRLS 2001	0.554	..
Uganda	SACMEQ II 2002	0.57	0.65
Singapore	PIRLS 2001	0.586	..
Namibia	SACMEQ II 2002	0.60	0.53
Namibia	SACMEQ I 1995	0.65	..
South Africa	SACMEQ II 2002	0.70	0.64
South Africa / Western Cape (this study)	Western Cape Primary School Pupil Survey 2003	0.72	0.44

Source: Postlethwaite 2004: Tables 3.6 and 3.7; and this study

Table X further shows that almost half the variance in test scores between schools can be explained by the HLM models (44% and 47%), but far less of the individual variance (24% and 14%), which may reflect the fact that individual ability and motivation cannot be captured in the observed variables.

Both the numeracy and the literacy models in the HLM model in Table X include a school level (Level 2) random effect on the intercept and on homework. The random effects on homework indicate that returns to homework are dependent on the school, with the mean slope coefficient that can either double or turn negative if the effect of one standard deviation change on the random effects is added or subtracted. Table X considers the interaction between school and household poverty and effort. It shows that moving from the bottom to the top end of the effort scale (the equivalent of starting to do homework more than three times a week when you previously never did homework) is expected to earn you just more than seven additional marks in the literacy test. It is telling that the model predicts that the same pupil could see a comparable rise in marks (5 marks) if he or she continues to do no homework at all, but moves to a more upmarket neighbourhood and attend a wealthier school. It is instructive that high effort has almost twice the payoff in richer schools (almost 13 marks) compared to poorer schools (7 marks), pointing to the limitations of poorer schools in assisting pupils to escape their socio-economic circumstances.

TABLE X: Expected literacy score of pupil by effort level, endowment of school and household income		
Attending poorer schools	**Low effort**	**High effort**
Pupil from poor household	43.6	50.7
Pupil from rich household	46.8	53.9
Attending richer schools	**Low effort**	**High effort**
Pupil from poor household	45.4	58.0
Pupil from rich household	48.7	61.2

NOTE: Poor and rich households were selected to be those with scores of 10 and –10 respectively (i.e. 10 percentage points above and below the grand mean of the index). Poor and rich schools were taken to be those with a score of 0.25 and -0.20 (above and below the grand mean for the index). Low effort was equated with a score of 1 on the homework frequency question, indicating pupils who reported never doing any homework. High effort was seen as a 4, a score associated with doing homework 4 times a week.

Appendix F: Data reliability

The survey appears to be reasonably reliable as conclusions from the analysis of this survey is broadly in agreement with findings in the empirical literature. A handful of survey questions were asked to several of the school representatives interviewed. The repeated observations can be used as a consistency check to gauge the reliability of the data. The table below shows how the management, the head of department and a teacher answered the question "Does the school distribute textbooks to individual pupils to keep?"

The table below shows how the management, the head of department and a teacher answered the question "Does the school distribute textbooks to individual pupils to keep?"

TABLE 1: Does the school distribute textbooks to individuals to keep?

Number of schools	Proportion of schools	Management interview	Head of Department interview	Teacher interview
14	21%	Yes	Yes	Yes
5	7%	Yes	Yes	No
5	7%	Yes	No	Yes
9	13%	Yes	No	No
1	1%	No	Yes	Yes
5	7%	No	Yes	No
7	10%	No	No	Yes
22	32%	No	No	No

Answers given by staff members are contradictory for 31 of the 68 school listed above. In spite of the noise, the variable appears to be capturing something of value. The average of the three repeated observations is positively and significantly correlated with the observed presence of a mathematics textbook.

Appendix E: Is there evidence of bias in the survey sample?

Despite prudent research design, there were a number of problems in the final data sets:

- Due to problems experienced in the field, some sampled schools or their replacement schools could not be tested. Three schools refused to participate in the survey and in six other schools testing of pupils could not be concluded in time. Replacements were often not possible due to the encroaching end of year examination period. This reduced the sample of schools with pupil test scores to 81.
- On the individual level, we do not have Census enumerator area information for 1394 of the total 2678 pupils (either because no address was given or otherwise because the given address could not be identified and matched to a Census enumerator area). These observations are discarded in regression analysis, leaving only 1284 of the initial sample of 2678 pupils.
- There are some missing values due to non-response in the school management module. Missing values are considerably higher in the classroom observation module where variables were imputed from a free-format fieldworker observation of classroom practices.

Analysis shows that there is little evidence of bias resulting from losing nine schools from our sample due to missing test scores. The performance of the nine schools that were dropped were statistically somewhat better than average on the Grade 3 tests earlier conducted by the WCED, as indicated by the low statistical significance of Bartlett’s test for equal variances between the included and the excluded groups. However, a chi-square test showed a statistically similar distribution (at the 99% level of significance) across the three sampling categories of the original sample compared to the 81 schools retained in sample tested. The similarity is evident from Table 1.

TABLE 1: Distribution of schools in original sample vs. schools for which pupil test data is available by relative performance, language of instruction and former department

		Distribution for schools in original sample				Distribution for regression under consideration			
		Afrikaans	English	Xhosa	Total	Afrikaans	English	Xhosa	Total
Poor performers	CED	2.2%	2.2%		4.4%	2.3%	2.3%		4.7%
	DET			4.4%	4.4%			4.7%	4.7%
	HOR	20.0%	3.3%	0.0%	23.3%	20.9%	3.5%		24.4%
	Total	**22.2%**	**5.6%**	**4.4%**	**32.2%**	**23.3%**	**5.8%**	**4.7%**	**33.7%**
Moderate performers	CED	4.4%	7.8%		12.2%	4.7%	7.0%		11.6%
	DET		1.1%	3.3%	4.4%	0.0%	1.2%	3.5%	4.7%
	HOR	11.1%	5.6%		16.7%	11.6%	5.8%		17.4%
	Total	**15.6%**	**14.4%**	**3.3%**	**33.3%**	**16.3%**	**14.0%**	**3.5%**	**33.7%**
Good performers	CED	3.3%	2.2%		5.6%	3.5%	2.3%		5.8%
	DET			3.3%	3.3%			3.5%	3.5%
	HOR	15.6%	10.0%		25.6%	14.0%	9.3%		23.3%
	Total	**18.9%**	**12.2%**	**3.3%**	**34.4%**	**17.4%**	**11.6%**	**3.5%**	**32.6%**
Total	CED	10.0%	12.2%		22.2%	10.5%	11.6%		22.1%
	DET		1.1%	11.1%	12.2%		1.2%	11.6%	12.8%
	HOR	46.7%	18.9%		65.6%	46.5%	18.6%		65.1%
	All	**56.7%**	**32.2%**	**11.1%**	**100.0%**	**57.0%**	**31.4%**	**11.6%**	**100.0%**

For descriptive analysis all available observations were used, resulting in fluctuating samples. The same strategy was followed with multivariate analyses, but here the average sample size was considerably lower because the resulting sample is one for which all variables in the model are available. To maximise sample size for the multivariate analysis, variables with more than 20 missing values were dropped from the sample.

The larger reduction in sample size for the multivariate analysis could also introduce bias. Sample sizes for the regression analysis mostly ranged between 45 and 60 schools, due to missing values on some variables included in the regressions. The impact of this reduction is considered by examining the sample retained in a reasonably robust regression of mathematical test scores. A chi-squared test showed that the statistical distribution of the 54 schools retained in the regression was similar across the three sampling categories at the 99% level of significance, whilst analysis of variance between the initial and resulting sample showed a 99.9% probability of equal variances of the earlier Grade 3 results, using Bartlett’s test. Table X compares the proportions of the initial sample to that of the regression sample under consideration. A startlingly similar distribution is observed between the two tables across the three sampling categories.

TABLE X: Distribution of schools in original sample vs. schools retained in regression under consideration by relative performance, language of instruction and former department

		Distribution for schools in original sample				Distribution for regression under consideration			
		Afrikaans	**English**	**Xhosa**	**Total**	**Afrikaans**	**English**	**Xhosa**	**Total**
Poor performers	**CED**	2.2%	2.2%		4.4%	1.9%	1.9%		3.7%
	DET			4.4%	4.4%			7.4%	7.4%
	HOR	20.0%	3.3%	0.0%	23.3%	22.2%	0.0%		22.2%
	Total	**22.2%**	**5.6%**	**4.4%**	**32.2%**	**24.1%**	**1.9%**	**7.4%**	**33.3%**
Moderate performers	**CED**	4.4%	7.8%		12.2%	1.9%	7.4%		9.3%
	DET		1.1%	3.3%	4.4%			1.9%	1.9%
	HOR	11.1%	5.6%		16.7%	9.3%	7.4%		16.7%
	Total	**15.6%**	**14.4%**	**3.3%**	**33.3%**	**11.1%**	**14.8%**	**1.9%**	**27.8%**
Good performers	**CED**	3.3%	2.2%		5.6%	5.6%	1.9%		7.4%
	DET			3.3%	3.3%			3.7%	3.7%
	HOR	15.6%	10.0%		25.6%	14.8%	13.0%		27.8%
	Total	**18.9%**	**12.2%**	**3.3%**	**34.4%**	**20.4%**	**14.8%**	**3.7%**	**38.9%**
Total	**CED**	10.0%	12.2%		22.2%	9.3%	11.1%		20.4%
	DET		1.1%	11.1%	12.2%			13.0%	13.0%
	HOR	46.7%	18.9%		65.6%	46.3%	20.4%		66.7%
	All	**56.7%**	**32.2%**	**11.1%**	**100.0%**	**55.6%**	**31.5%**	**13.0%**	**100.0%**

Thus, for the particular regression sample under consideration, non-response and missing values did not introduce significant bias in terms of representivity of schools. Of course this does not prove that there could not be bias present in other regression samples, but it does provide some comfort. In the multivariate analysis the school level results will consequently be regarded as representative of schools in the five EMDCs covered.

Appendix G: Clustering among observations

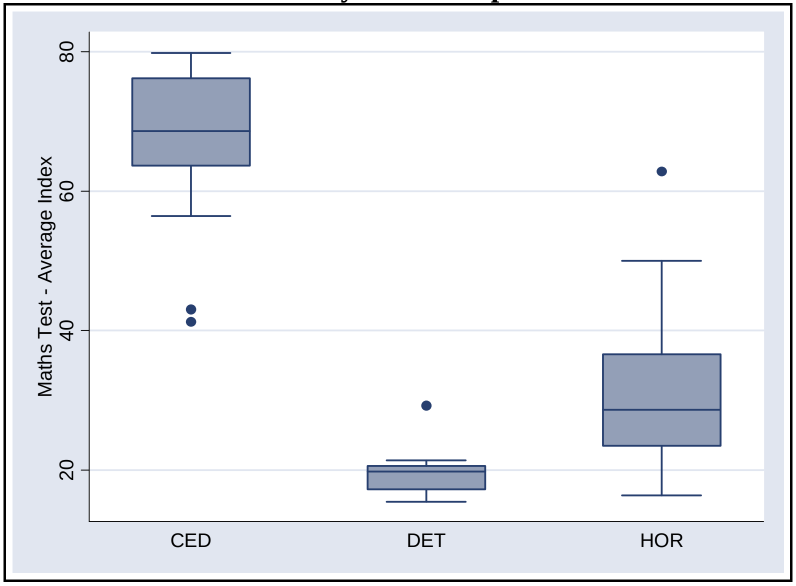
Statistical analysis can be complicated by the clustering of observations. The cross-tabulations in the statistical appendix show that the influence of historical factors endures. We find that schools from the same former department are often assigned to the same variable categories. Cluster analysis can be used to illustrate the problem. Using the Euclidean distance as the measure of dissimilarity, the observations are partitioned into four non-overlapping groups or types according to dissimilarities in numeracy and literacy test results, school fees and two variables related to assessment. A cross-tabulation of the resulting school types shows the considerable overlap between these estimated cluster types and former department. For instance we see that school type 4 consists entirely of ex-HOR schools. Also, 9 of the 10 ex-CED schools are defined as school type 2 and 9 of the 10 schools defined as school type 2 are ex-CED.

TABLE X: Cluster school types vs. former department

Cluster school type	Former department			Total
	CED	DET	HOR	
1	0	4	24	28
2	9	0	1	10
3	1	3	6	10
4	0	0	6	6
Total	10	7	37	54

The box-and-whiskers plot below of numeracy tests scores by former department shows the polarisation associated with the ex-department clusters. The mean numeracy test scores for ex-CED, ex-HOR and ex-DET schools are 67.0, 30.7 and 19.8 respectively.

FIGURE X: Box-and-whiskers plots of average mathematics test scores by former department



The detected clustering can complicate statistical analysis in at least two ways:

- It can be difficult to distinguish the separate influences on the education process because variables are highly correlated with one another and also with key variables such as poverty, language group and ex-department.
- If variable values are not only highly correlated within the cluster, but also polarized across clusters, then the combination of various clusters in one regression can lead to spurious regressions.

To test for the presence of the statistical problems associated with clustering, additional models are estimated using only ex-HOR schools as a sample. There are not enough schools in the other two ex-department groups to allow estimation as a separate sample. These regressions are referred to in the text but not reported separately in the tables.

Appendix H: Theoretical Model

The empirical literature on school quality in South Africa is weak on casual factors relating to classroom instruction and school management mainly because up to now local data sets have not contained much information about these factors. This report exploits the richness of the 2003 Western Cape Primary School Progress survey to learn more about the impact of different factors concerning classroom instruction and school management. This richness potentially allows us to move beyond the proxy of former department to investigate the factors that explain the differences in schools performance.

The report takes the model proposed by Hanushek (2002) as a starting point. He suggested the following specification for an education production function:

$$y = \beta_1 F \quad + \beta_2 P \quad + \beta_3 S \quad + \beta_4 A + \mu$$

where y represents the performance of pupil *i* at time *t*

F is family inputs cumulative to time *t*

P is cumulative peer inputs to time *t*

S represents cumulative school inputs to time *t*

A is the pupil's individual characteristics, including innate ability and

μ is an error term

ßs are the coefficients (returns to inputs and characteristics).

Adapting this theoretical model for local conditions, an indicator of the proficiency in the language of instruction is added to the original model. The causal process could thus be expressed as:

$$y = \beta_1 F + \beta_2 P + \beta_3 S + \beta_4 L + \beta_5 A + \mu$$

where L indicates the pupil's proficiency in the language of instruction and all other variables are defined as in the previous model

In the models above the acquisition of education is described as a production function. This application of the production function approach to education is often criticized because a production function assumes clearly distinguishable inputs and optimal efficiency in the translation of inputs to outputs. In the education process, it is often difficult to disentangle different inputs. For instance, in the Western Cape we find that poverty is substantially higher among Xhosa-speakers, which makes it difficult to detect the separate influences of socio-economic circumstances and proficiency in the language of instruction. If the school management variables are interpreted as efficiency indicators, it could be argued that this efficiency assumption is no longer applicable. Despite its shortcomings, the education production function approach has become a standard tool for analysing the effect of different factors on education outcomes.

The focus of this report is on examining the contribution of school inputs and processes – and within this group of variables specifically the impact of classroom instruction and school management – on the numeracy and literacy of pupils.

Appendix I: Methodology for estimating model

Based on the empirical literature on quality education outcomes in South Africa, apartheid-era factors (represented by former department and the predominant language spoken by the pupils in the school) are added to the list of variables. Although previous work has not identified gender as an important determinant of education attainment or quality in South Africa, it is included in the list as a standard control.

In the selection of appropriate variables to estimate the defined model, we encounter a common problem. Our model requires that we should control for the pupil's ability to understand and learn, but the survey did not measure the ability of pupils. Information on pupil ability is often not available in surveys. Excluding ability from the model can bias all the variable coefficients if one or more of the variables included in the model are correlated with ability. The standard example is that the income and educational attainment of the parents will be correlated with the ability of the pupil, but due to the influence of apartheid, there is not much reason to suspect a strong correlation of cognitive ability with the socio-economic indicators.

Another problem with cross-section data sets is that surveys usually provide information on school management, classroom practices and family circumstances at a point in time, while the available school quality output variable that the analysis attempts to explain is often a test score that measuring learning that has been accumulated over time. When using point-in-time estimates to represent cumulative variables the implicit assumption is then that the current state is correlated to previous states. This assumption is not unreasonable as many of these factors are institutional or social and these forces are slow-changing. School management variables are presumed to generally be more slow-changing than the year-to-year classroom circumstances and practices and are thus expected to yield stronger results with the cumulative school quality variables.

The school level model is the result of an iterative elimination process. Due to the extremely high correlation of the school management and classroom variables with former department and the overpowering influence of the latter variable in regression analysis, ex-department is initially excluded from the model estimation process to focus on the mechanisms through which advantage and privilege work. The richness of the management and classroom variables in this data set reduces the likelihood of serious omitted variable bias.

The model selection process starts with a model including one or two variables from each category. In successive rounds the variable with the least significant relationship with the numeracy and literacy test scores were eliminated and the model was re-estimated. This process continued until all variables left in the model had a significant relationship (p-value lower than 0.1) with the numeracy or literacy test scores.

To ensure consistency with the specified model, care was taken that all categories were represented in each of the models estimated. Within categories variables were ranked according to the significance of the relationship between the variable and the numeracy and literacy test results.

A similar process was followed for the selection of the individual level models, only the process was simpler. Coefficients are more stable because of the considerably larger sample. Also, the model selection process is less complicated because there are fewer candidate variables available.

After an acceptable model had been identified through this process, the model was subjected to repeated testing to ensure stability. One-at-a-time, all the variables not included in this model were added to the model. To investigate the hypothesis that the regression could be spurious because it compares schools from different systems, the model was re-estimated limiting the sample to contain only ex-HOD schools.

As there was some evidence of heteroskedasticity, all regressions show Hubert/White robust standards errors. We also test for the impact of outlier values.

The third model combines the individual and school level models to look at both between school and within school variation in test results. Here a two-level hierarchical linear model is used. Hierarchical linear models are an improvement on both individual and school level models because they avoid the perils of aggregation at the school level by modelling variation on the individual level and prevent the underestimation of the error terms by taking account of the nested/grouped structure of the individual data. Information about higher levels, such as classroom and school, can be used to predict the slopes and intercept parameters of variables in lower levels in the model (e.g. individual level). As a starting point for the hierarchical model, the predictors from the previous two models (individual and school level) are used.