
USING THE HIERARCHICAL LINEAR MODEL TO UNDERSTAND SCHOOL PRODUCTION IN SOUTH AFRICA

MARTIN GUSTAFSSON

Stellenbosch Economic Working Papers: 01/07

KEYWORDS: EDUCATIONAL QUALITY, EDUCATION POLICY, EDUCATION RESOURCES, SACMEQ, SOUTH
AFRICA
JEL: I21, H52

MARTIN GUSTAFSSON

E-MAIL: martingust@worldonline.co.za



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

USING THE HIERARCHICAL LINEAR MODEL TO UNDERSTAND SCHOOL PRODUCTION IN SOUTH AFRICA

MARTIN GUSTAFSSON¹

ABSTRACT

The emphasis placed in the existing South African school production function literature on better skilled teachers and better school management is discussed. Ordinary least squares and hierarchical linear production function models, using 2000 SACMEQ data, for the country and for a sub-set of historically disadvantaged schools, are constructed. Ways of making the results more readable for policymakers are explored. The importance of physical infrastructure, textbook and nutrition budgets is highlighted by the models. Correct allocation of teaching and management time in schools, less learner repetition, and better teaching methodologies stand out as important school and classroom management imperatives.

Keywords: Educational quality, Education policy, Education resources, SACMEQ, South Africa
JEL codes: I21, H52

¹ Economist, Research Triangle Institute. Based at Department of Education, Tshwane (Pretoria). Most of the research in the paper was undertaken whilst the author was a student at the University of Stellenbosch (2004-2006).

USING THE HIERARCHICAL LINEAR MODEL TO UNDERSTAND SCHOOL PRODUCTION IN SOUTH AFRICA

1 INTRODUCTION

Achieving expenditure equity in South Africa's public schooling system, a system that encompasses more than 95% of all schooling in the country, has justifiably been a central concern of the post-apartheid government, and, barring some residual issues, equal public spending across all learners has been achieved (Gustafsson and Patel, 2006). Policy attention is increasingly focussing on what is by now clear evidence of unacceptably low levels of learner performance across the bulk of historically disadvantaged schools, but even, by international standards, poor quality in historically advantaged schools. In a context of rising public expenditure in real terms in the schooling system (a real increase of around 32% between 2002 and 2008), understanding the relationship between publicly funded inputs and quality schooling becomes critical. Education production functions can provide important guidance in this regard. This article contributes to the South African (and developing country) production function discourse through analysis of 2000 SACMEQ data pertaining to South Africa. The article draws from, and expands upon, a Masters thesis and a related paper presented at a SACMEQ conference (Gustafsson, 2005 and 2006).

2 THE EDUCATION PRODUCTION FUNCTION POLICY DEBATE

In the quest for social welfare improvements, education cost benefit analyses are useful in that they demonstrate the returns, in the form of improved income for the individual and economic growth for society, gained from investments by the individual and by society in education. For the policymaker, these analyses can assist in optimising the spread of investments across the different types and levels of education. However, they are crude in the sense that they compare money and time spent against income gains. They do not permit a view into the black box of internal education efficiency, or the relationship between education resources and learning. This is obviously a key concern for education policymakers where difficult resourcing choices must be made. Education economists have tended to turn to an estimation of the education production function to understand the inside of the black box. This function, which we can represent as

$$Y_i = F(X_{1i} \dots X_{ni}, Z)$$

must include three things. Firstly, it must include a learner performance variable, Y , which is typically the mathematics or language score from a standardised test. Secondly, it must include a number of education resources, the X values. Often the interpretation is very broad, so for instance teacher level of satisfaction could well be considered an education resource. Thirdly, the function must view the effect of the education resources as conditional on the level of socio-economic status (SES) of learners (Z). This is because of the clearly proven positive correlation between SES and performance, independent of the education resources available. Typically, Z will include measures of both material welfare and the level of education of the learner's parents. The level of the analysis, or i , may be the individual learner or the school.

There is considerable and probably growing interest in the education production function approach in a number of quarters. Academic articles using this approach are common, the South African Department of Education is interested in it (South Africa, 2004: 45), and it receives attention in the Education for All reports (UNESCO, 2005: 65). There is not universal acceptance, however. Adherents of the 'effective schools approach', for instance, would advocate a less statistical analysis of just a few, well-performing schools, and working backward, or inductively, to identify the education resources and practices that can be associated with high performance. Clearly, the two approaches can be considered complementary (Monk, 1990: 413).

Education authorities are often interested in ranking schools on the basis of performance. In a typical 'league table', an output measure such as the school average of learner scores in a test is used to rank schools. The production function provides a useful basis for an alternative ranking system that is in many ways superior to the typical league table approach insofar as this alternative takes into account the differing conditions under which schools produce education, both in terms of education resources and learner SES. Crouch and Mabogoane (1998) put forward a methodology for such an alternative that was used in a very public manner in South Africa. Schools may also be ranked according to a formal definition of institutional efficiency. Johnes (2004: 648) advocates the

use of data envelopment analysis (DEA), a non-statistical linear programming approach, to establish ‘efficiency boundaries’ against which the output of individual institutions can be compared in order to calculate institution-specific efficiency scores.

Often the scepticism expressed about the findings of individual production function studies relates to the fact that these studies generally use datasets that were not compiled expressly to cater for such analysis. For instance, the SACMEQ data on which this study is based includes background variables on schools, teachers and learners which were identified in a fairly general manner, on the basis of what Ministry officials across the thirteen countries wanted to know about participating schools. Whether these variables were thematically and structurally suitable for a production function analysis was not a primary concern. The datasets used in production function studies frequently omit certain education resource variables that are obviously important. For instance, the paucity of information on school management resources and practices in the SACMEQ dataset seriously limits its production function potential. As a result of these kinds of limitations, analysts have become increasingly hesitant to draw hard policy conclusions from individual studies, and have turned to offering policy advice on the basis of patterns that are repeated across the findings of many studies. A few formal methodologies for this kind of meta-analysis have been proposed (Lee and Barro, 2000: 8). The present study needs to be understood in this context. It would be unjustifiable to draw hard policy conclusions from just this analysis without also considering other, similar studies.

One meta-analysis is that of Hanushek (2002), which is partially reproduced in UNESCO’s 2005 Education for All *Global Monitoring Report* (UNESCO, 2005: 65). This meta-analysis, which covers 96 studies dealing with developing countries, tells us that four things matter above all for learner performance: teacher training, teacher experience, spending on non-personnel inputs, and school physical infrastructure. It furthermore tells us that two things do not clearly matter: teacher salary and learner/teacher ratio.

UNESCO and Hanushek differ rather fundamentally in their conclusions, though they use the same data (Gustafsson, 2006: 51). Hanushek (whose 2002 title is *The failure of input-based schooling policies*) has over the years argued strongly to policymakers that they tend to worry too much about the budget split across various inputs, and education resources that cost additional money, as opposed to improving the productivity of the existing resource bundle in schools. Hanushek has tended to use his own production function studies to demonstrate that marginal increases in inputs do not make a difference (here we should understand inputs to exclude management practices). Hanushek’s argument is relatively solid when it comes to developed countries, where education resourcing levels are high enough to render the marginal product of inputs low. However, the argument is weak when made in the context of poorly resourced schooling systems in developing countries. In fact, Hanushek and Luque (2003) make this argument with respect to developing countries on the basis of just five developing countries, two of which have a high human development index (UNDP definitions used).

An obvious problem with the meta-analysis approach is that it glosses over possible production function differences with respect to, firstly, individual countries and, secondly, the level of the schooling system (primary or secondary). In South Africa, a small set of studies has emerged over the last decade. Case and Deaton (1999) counter the usual production function findings by arguing that smaller classes do lead to improved learner performance. Importantly, their analysis is based on the situation in black African schools in 1993, before a major equalisation of class sizes was introduced. Crouch and Mabogoane (1998) (who examine secondary schools) lean towards the Hanushek argument when they argue that school management factors which are not reflected in their dataset ought to be a key concern for the policymaker (their title is *When the residuals matter more than the coefficients*). Crouch and Mabogoane furthermore question the statistical significance of Case and Deaton’s findings. Crouch and Perry (2002) (who look at primary schools) confirm the general finding that teacher training matters, though they express this variable in terms of the results obtained by teachers in a subject content test. Van der Berg and Burger (2003) find teacher salary to be a significant predictor of performance (in secondary schools), but emphasise that they consider salary to be a proxy for the teacher’s level of qualification. Van der Berg (2005), using the same SACMEQ dataset as is used in this study, also leans towards the importance of improving the management of inputs, as opposed to increasing the inputs themselves. Bhorat and Oosthuizen (2006), focussing on Grade 12 performance, conclude that certain infrastructure inputs matter, for instance the quality of classrooms, and the existence of libraries and computers, that the learner/teacher ratio exerts an insignificant effect, that the apartheid ex-department of the school is significant and that teacher quality (as measured by proxy variables) is important. For now, having better trained teachers in schools and improving school management stand out as the common threads emerging from the South African production function literature.

3 THE DATA AND THE MODELLING APPROACH

The data used for the current study are the South African portion of the data collected in 2000 by the Southern and Eastern African Consortium for Measuring Educational Quality (SACMEQ)². SACMEQ was first run in 1996, again in 2000, and the next run is planned for 2007. SACMEQ involves administering standardised reading and numeracy tests to a representative sample of primary school learners in thirteen countries. School principals, teachers and learners complete background questionnaires on which the input variables of this study are based. In 2000, SACMEQ covered Grade 6 learners. In the case of South Africa, the dataset covers 3,163 learners from 169 schools. Weights attached to schools make the data representative of the population of all Grade 6 learners in the country.

SACMEQ can be viewed as part of a growing presence of cross-country monitoring programmes examining learner performance and its determinants. At the global level, the TIMSS programme has increasingly attracted participation from developing countries – from six such countries in 1995 to 23 in its last run in 2003. A regional equivalent to SACMEQ in Latin America is the Laboratorio programme. Participation in these kinds of programmes and engagement with the results should be regarded as important for policymakers.

For the purposes of this study, a policy framework made up of 22 policy areas was devised. From the 381 variables of the three background questionnaires, it was possible to extract 18 variables dealing with 18 of the 22 policy areas. Both relevance in South African policy terms, as well as correlations with test scores and correlations between input variables were taken into account in the determination of the 18 variables. Of the new 18 variables, six were derived from single variables in the original dataset, whilst 12 were constructed from more than one original variable. In addition to the 18 new variables, three other variables were regarded as important: the learner's age, the learner's gender, and whether the school principal thought teachers frequently arrived late at school. The last of these three stood out for its high correlation with the test scores (this was true for South Africa, but not for the other SACMEQ countries). In total then, 21 input variables were considered to be available for the input-output analysis. These variables are listed in Tables 1 to 3, either as included or as excluded variables, and definitions appear in Table 4.

To explain the structure of the hierarchical linear model (HLM), it is useful to first consider the usual one-level ordinary least squares (OLS) model:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{\beta}_4 X_{4i} + \hat{u}_i$$

Partly in line with what is actually reported in Tables 1 and 2 below, we can think of X_1 as the level of school infrastructure, X_2 as the learner's age, X_3 as the value attached to the teacher's classroom methodologies and X_4 as the learner's SES. The HLM might be structured as follows (this is a shortened version of the model presented in Table 2):

$$Y_{ij} = (\hat{\alpha}_{00} + \hat{\alpha}_{01} X_{1j} + \hat{\epsilon}_{0j}) + \hat{\beta}_2 X_{2ij} + (\hat{\alpha}_{40} + \hat{\alpha}_{41} X_{3j}) X_{4ij} + \hat{u}_{ij}$$

The model describes the production function for learner i in school j . Here the school infrastructure variable, a school-level or level 2 variable, becomes nested under the learner-level, or level 1, intercept β_0 . This allows the level 1 intercept to differ from school to school. In addition, the level 1 slope coefficient for learner SES will differ from school to school, because the level 2 variable value of classroom methodology has become nested within the slope coefficient β_4 . Three benefits to the HLM structure stand out (Raudenbush and Bryk, 2002: 7). Firstly, because there are now separate error terms for levels 1 and 2 (ϵ and u), it is possible to partition the variance across the two levels. Put differently, between-school and within-school inequality can be separated out in the model. Secondly, the model allows for an analysis of the interaction between variables from different levels. For example, the model can indicate how the value of classroom methodology impacts on the ability of learner SES to be translated into learner performance. Thirdly, the model permits the analyst to examine school-specific coefficients, which can yield important information about individual institutions. This study will take advantage of the first two benefits.

² Permission to use this data was obtained from the Harare SACMEQ office. Support received from Paul Murimba and others in the SACMEQ team in accessing and understanding the dataset is appreciated. SACMEQ's website address is www.sacmeq.org.

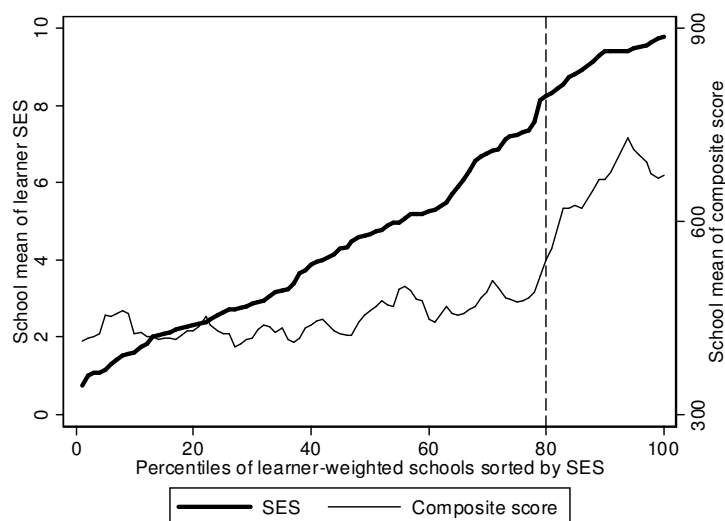
In the HLM analysis, the decision was taken to average teacher background values at the school level, as opposed to constructing a three-level HLM. This was because in the great majority of schools (153 of 169) there was only one teacher for each subject in each school, so no averaging was necessary.

In both the OLS and HLM analyses, a backward stepwise approach was used to exclude variables. In the case of the OLS model, this approach was partly automated, and partly manual. In the case of the HLM, the software used (HLM 6.02a by Raudenbush, Bryk and Congdon) did not allow for automated stepwise exclusion, or for the automated testing of the many possible combinations of level 1 and level 2 variables, and consequently the exclusion and testing process had to be entirely manual. Policy relevance, overall unexplained variance and the significance of individual coefficients (using the *t* statistic) were used to determine the final structure of all the models. In the HLMs, a less stringent requirement for the retention of level 2 variables relative to level 1 variables was followed, in line with the recommendation of Raudenbush and Bryk (2002: 268), who specify a *t* statistic of absolute value 1.0 as a threshold for possible inclusion of a level 2 variable.

The HLM software still requires improvement, and is currently rather cumbersome to use. Moreover, conceptually the HLM is considerably more complex than the one-level OLS model, partly because of the Bayesian methodology used to estimate coefficients and the variance statistics. Johnes (2004: 647) describes the multi-level model (or HLM) as ‘computationally intractable’.

Models were generated not just for the full South Africa set, but also for a sub-set which excluded the most advantaged schools in terms of the school mean of SES. South Africa (together with Namibia and Uganda) displayed a distribution of test scores at the learner level which appeared bimodal, suggesting the existence of, in a sense, two schooling systems within the country. This would be consistent with South Africa’s apartheid history, and also the dual economy concept in the economics of development theory. To construct the sub-set, the 20 per cent of weighted learners in the sample attending schools with the highest school means of the learner SES variable were excluded. Figure 1 indicates that though the distribution of learner SES (as calculated from the SACMEQ dataset using a factor analysis approach) was relatively linear, excluding the top 20 per cent according to SES resulted in the exclusion of a portion of the sample with considerably better performance on average (the distribution of performance was clearly not linear). The excluded 20 per cent would encompass more than learners in historically white schools – only around 8 per cent of Grade 6 learners in the public schooling system were in such schools in 2000. In fact, some analysis of the 2004 Grade 12 Senior Certificate results reveals that historically white, African and Indian schools account for 46, 34 and 13 per cent respectively of the ‘performance elite’, if we consider this to be Grade 12 learner from the one-fifth of schools with the best examination pass rates (own calculations using Department of Education data). However, though historical distinctions may have become somewhat blurred at the top performance end, at least 85 per cent of the schools to the left of the 80th percentile in Figure 1 can be assumed to be historically African schools.

Figure 1: Distribution of SES and performance



Source:: SACMEQ, 2000. The ‘composite score’ is the mean of the reading and mathematics scores.

4 RESULTS

Of the 3,163 learners and 169 schools in the full South Africa SACMEQ dataset, 2,612 learners and 140 schools were considered part of the historically disadvantaged (HD) sub-set. Instead of imputing missing values, observations with missing data were excluded from the analyses that produced the results in Tables 1 to 3. In no analysis were more than 5% of observations excluded. Table 1 describes the OLS reading model for historically disadvantaged schools, whilst Table 2 describes the HLM reading model for the same schools. The results in this format for a further six models are not provided here, though all eight models are used in Table 3, which, it will be argued, is more accessible for the policymaker.

The partitioning of the overall variance across the two levels (between-learner and between-school) in Table 2 illustrates part of the reason why it is important to perform a separate analysis of the HD sub-set of observations in the case of South Africa. In the HLM null model for South Africa, or

$$Y_{ij} = \hat{\alpha}_{00} + \hat{\epsilon}_{0j} + \hat{u}_{ij}$$

just over two-thirds of the variance in scores, whether we consider reading or mathematics, is between schools, as opposed to between learners. This is an unusual pattern if we make an international comparison. In typical country datasets, most variance by far is between learners (Gustafsson, 2006: 144). This is the case in most of the SACMEQ countries – only South Africa, Namibia and Uganda display more between-school variance than within-school variance. Surprisingly, Brazil, a country often equated with South Africa in terms of income inequalities, displays the typical pattern of more between learner performance inequality than between school inequality, both at the Grade 4 level and with respect to 15 year olds (Willms and Somers, 2001; Own analysis of PISA data available at www.oecd.org). The typical pattern is obtained with the South African data, if historically advantaged and historically disadvantaged schools are considered separately, underlining how, in a sense, one is dealing with more than one schooling system in the case of South Africa. The variance statistics for the null reading model for HD schools indicate that between-learner variance constitutes around 54% (3861 over 7197 – see Table 2) of the total variance – a value over 50% is what one would expect from most schooling systems.

In the Table 2 model, the level 1 variables are able to explain only 10% of the between-learner variance, whilst the level 2 variables are able to explain 55% of between-school variance. The greater explanatory power of the level 2 variables characterises practically all HLM models dealing with the schooling process. That this should be so is partly because questionnaires are largely designed to detect between-school and between-teacher differences, but also an indication of the strength of peer effects in schooling.

It is difficult to present the raw results of an HLM, such as that in Table 2, in such a way that they are meaningful to policymakers. Table 3 was devised to provide a more normalised and readable representation of the Tables 1 and 2 results. Table 3 concentrates only on the impact of interventions in the HD sub-set of schools (this is commonly the focus in the policy dialogue). The actual performance of the 10th, 50th and 90th percentiles of weighted HD learners is given. Moreover, the predicted mean performance for HD learners using eight different models is given. The ‘HLM all’ models are the HLM models with coefficients produced using data from all South African schools (not just HD schools), though the mean score improvements referred to in Table 3 are those of HD school learners only. The ‘HLM HD’ models are the Table 2 model plus the corresponding model for mathematics – here only the HD schools data were used to generate the coefficients. Four OLS models, namely the Table 1 model, the corresponding mathematics model, and a further reading and mathematics model where coefficients were generated using data from all schools, were also used in Table 3. The eight predicted means are very close to the actual weighted mean scores for HD schools, which are 449 for both reading and mathematics (this is not reflected in Table 3). A look at the impact of the pre-service training variable explains the logic of the table. This variable refers to the total years of education of the teacher. With observations sorted according to level of pre-service training, the value at the 10th percentile (considering only HD schools) is 14 years, whilst the values at the 50th and 90th percentiles are 15 years. In the case of this variable, these values are the same for reading and mathematics – they could differ, however, for variables derived from the educator questionnaire. To simulate a policy intervention where the years of training of teachers was improved, the values for the 1st to the 89th percentiles were set to the value of the 90th percentile, in other words 15 years. The adjusted data were run through the eight models in order to gauge the new predicted mean score for HD schools. For the ‘HLM all’ model, the predicted mean reading score rose from 451 to 459, in other words by 8 points. The ‘OLS all’ model produced a gain of 12 points with respect to the reading score of HD learners. And so on. In the case of the ‘HLM HD’ model, the cells are blank because the pre-service training variable was not deemed sufficiently

significant to be included. A variable such as the teaching load of the principal, which has a negative association with performance, had to be treated in a reverse manner. Here values had to be adjusted *downwards* to 0 hours per week in order to simulate the performance impact of the change. Table 3 thus allows us to examine the predicted performance changes implied by many models in a standardised fashion, with direct reference to a statistic that is easily understandable, namely the mean score of historically disadvantaged learners. Arguably, this is more readable presentation of production function data for the layman compared to the more commonly used formats of Tables 1 and 2.

The HLM and OLS results display notable differences. In the OLS models, variables are more often significant and, when significant, display a larger impact, relative to the corresponding HLMs. In 44 comparisons (based on Table 3) variables exert a greater impact in the OLS model, against just 7 comparisons where variables exert a greater impact in the HLM. The difference is particularly marked with respect to the home background variables, especially learner SES, where the OLS models reflect a much larger impact than the HLMs. In fact, such are the interactions between SES and other variables in the all-data HLM model for mathematics that whether the SES level of HD learners is raised to the 90th percentile or dropped to the 10th percentile, a negative impact results. This is an important finding. Either the HLM models are under-stating the magnitudes of the impacts, or the OLS models are over-stating them.

A second methodological question is whether constructing separate models based on just HD data yields different magnitudes of impact. Table 3 indicates that the separate HD-data models did bring to the fore patterns that were obscured in the all-data models, or render insignificant effects that had stood out in the all-data models. For example, an extra year of pre-service training appears to result in an increase of 8 or 6 performance points for HD learners in reading and mathematics when the all-data HLMs are used. However, in the HD-data HLMs, years of pre-service training exerts no significant impact at all. This difference in results is almost certainly attributable to the fact that not only were teachers in HD schools trained less under apartheid in terms of years, relative to teachers in historically advantaged schools, the training offered tended also to be of a poorer quality. The HD-data models also reduce the positive impact of being a girl. This may be attributable to larger pressures in historically disadvantaged households for girls to perform household work. Interestingly, in the HLMs the relative importance of SES and parent education switches when the analysis moves from the all-data models to the HD-data models. In the all-data HLMs parent education exerts the larger impact, whilst in the HD-data HLMs the impact of SES is always greater than or equal to the impact of parent education. This could be indicative of weaker benefits to years of schooling in the apartheid education system designed for black South Africans. We can conclude that the continuing differences between the historically advantaged and disadvantaged, in terms of not just performance but the dynamics of schooling, were still large enough (in 2000) to render any production function based on all the data suspect for the purposes of drawing policy conclusions. Some analysis that focuses exclusively on historically disadvantaged schools, and their specific dynamics, remains important for the policymaker.

School infrastructure stands out as a particularly powerful variable. Improving the quality of this from an index value of 0.2 to 6.6 is associated with an increase of 39 performance points in reading and 11 in mathematics. A value of 0.2 typically means none of the following existed in the school: toilets, library, school hall, staffroom, principal's office, storeroom, photocopier, computer, tuckshop. A value of around 6.6 typically means that all the items except for two are present – typically the missing ones are the school hall and the tuckshop. Crucially, a correlation matrix of the 21 input variables reveals that the highest correlation is that between the school infrastructure variable and the proximity to urban facilities variable – the correlation coefficient equals 0.78. Moreover, of all the 21 input variables, only school infrastructure and learner SES display a higher correlation with the reading score than the proximity to urban facilities variable. Clearly, the infrastructure variable is masking some rurality effects, for instance school remoteness, which in turn indicates attractiveness for teachers and hence teacher quality (an input that is rather weakly measured in the dataset, partly because South Africa was one of two SACMEQ countries that excluded the teacher testing component of the programme). Any use of the physical infrastructure improvement lever would need to be coupled with strategies to counteract the disadvantages typically associated with remoteness from urban facilities, in particular strategies to attract better teachers to remote schools.

The policy implications of the strong association between years repeated and performance in Table 3 (no repetition is associated with an improvement of 25 reading points in the HD-data model, for instance) are difficult to understand. In all the eight models, the school mean of years repeated is retained, suggesting strong links between the proportion of learners who repeat, and the number of times they repeat, on the one hand, and the performance of *all* learners on the other hand. Peer effects seems to be at play. Notably, exactly half of HD

school learners have never repeated. The repeater variable is different from the other input variables insofar as it is to a large degree the result of poor performance (in previous years). Yet it is interesting to note that it is not the input variable with the strongest correlation with performance – it displays only the 6th highest correlation with the reading score, for instance (at the level of the learner). To a large degree, repetition practices are localised at the school level, and system-wide approaches are not recognised or followed. There appears to be a need to understand best practices better. Eisemon (1997) and others warn against a simplistic adoption of automatic promotion, and stress that country-specific repeater dynamics must be well understood before firm policy positions can be taken. There is clearly a need for more dedicated research into the phenomenon of grade repetition in South Africa.

A number of the Table 3 impact statistics are of importance for teachers and school principals, and those who train them. The HD-data models suggest that better teaching methodologies, even in the absence of increases in levels of pre-service and in-service training, can result in score improvements of 5 in both subjects. The reading methodology index that predicted good performance best attaches a high value to the promotion of listening skills, and having parents sign homework books. In mathematics, allowing learners to work occasionally on their own and interacting on a one-on-one basis with individual learners emerged as especially valuable. There were few school management variables in the data, but it is noteworthy that the teaching load of the school principal (conceivably the inverse of the amount of time spent managing) is a significant predictor of reading performance in the HD-data models (OLS and HLM), though not in the all-data models. Also on the reading side, the school factor with the greatest impact after infrastructure and repetition, is the teacher's formal teaching hours in a year. Raising the formally timetabled teaching time to levels seen in the better HD schools would raise the average reading score by 7 points. The teacher latecoming variable, based on a question which asked the school principal about the frequency of this phenomenon, is interesting. In 96% of HD schools this was identified as a problem, but only in 36% of historically advantaged schools. Using the all-data model coefficients, and simulating the complete removal of the teacher latecoming problem from all HD schools, raises the reading and mathematics scores by 20 and 32 points respectively. This is arguably an example of an inappropriate statistic arising out of the use of a uniform model for a highly segmented system. Nevertheless, the fact that the problem is reported to exist in 96% of HD schools is telling. The poor performance of the intensity of evaluation directed towards teachers variable seems mostly to be an indication of a lack of good data on this input in the dataset.

Other Table 3 impact statistics are of more relevance to government policymakers. Nutrition emerges as a factor with a measurable impact on performance. Ensuring that children eat three meals a day is associated with an increase of 4 performance points in reading and 2 in mathematics. The textbooks per learner variable was capped at 0.5 (meaning two learners share one textbook) given that most of the textbook impact occurred in the range of zero to 0.5. This mirrors findings in other countries (Glewwe *et al*, 2000:3). The unadjusted data would have displayed a value of 1.0 at the 90th percentile, for both subjects. Just ensuring that the ratio reaches 0.5 for each learner raises the reading score by 4 points (in the HD-data HLM). That in-service training should display little to no impact in the HD-data models, despite the fact that practically all of the 70% of HD teachers who received this training classified it as useful, seems surprising. This could point towards quality problems in the training programmes. However, this could also be a result of the fact that training programmes are often deliberately targeted towards poorer performing schools. This, in turn, highlights the problems in using cross-sectional and not time-series data. In keeping with most of the literature, the class size variable appears not to be a prominent predictor of performance, this despite the fact that 19% of HD learners were in classes with more than 50 learners. The earlier point about the trade-off between the principal's teaching time and the quality of school management does however remain important from a teacher provisioning point of view. The intensity of district support variable is based on the number of visits by 'inspectors' during the previous three years, with proximity to a resource centre counting as an additional 10 points (just over a third of HD teachers did report having access to such a centre). Again, the general absence of any discernable impact in the HD-data models could point towards poor quality in the service, or could be a result of how district support is targeted across schools.

Table 1: OLS results (historically disadvantaged reading model)

R ² = 0.317		n = 2514	F = 77	level=0.95
dependent variable: reading score				
		coefficient	beta coeff.	t stat
	constant	225.967	4.0	
E	years pre-service training	14.092	3.9	0.066
E	class size squared	-0.005	-3.8	-0.073
E	class methodology value	4.853	6.7	0.119
E	teacher hours in a year	-0.021	-5.2	-0.098
S	level of school infrastructure	9.729	12.2	0.264
S	principal's teaching load	-0.724	-3.1	-0.059
S	intensity of district support	-0.429	-2.9	-0.051
S*	school avg. number of years repeated	-30.099	-6.6	-0.131
S*	school avg. learner is a girl	75.005	5.6	0.096
L	number of years repeated	-9.548	-5.4	-0.102
L	textbooks per learner	25.310	3.2	0.055
L	average number of meals per day	5.588	3.0	0.051
L	years of schooling of parents	1.442	4.6	0.083
L	learner SES	3.830	6.6	0.129
L	learner's age in years and months	-5.188	-5.7	-0.103

Excluded: days of in-service training; evaluation intensity; teacher SES; level of parent involvement; principal's years of pre-service training; proximity to urban facilities; degree of teacher latecoming. The first column indicates whether the variable level is the Educator, the School, or the Learner, with S* denoting the school mean of a learner-level variable.

Table 2: HLM results (*historically disadvantaged reading model*)

dependent variable: reading score		Level 1 units:	2514
		Level 2 units:	134
<u>Fixed effect:</u>		coefficient	t stat
	For intercept β_0		
	intercept	433.669	28.5
S	level of school infrastructure	10.362	4.9
L	For number of years repeated		
	intercept	-9.279	-5.0
L	For textbooks per learner		
	intercept	27.652	2.7
L	For average number of meals per day		
	intercept	10.502	3.0
E	teacher hours in a year	-0.005	-1.4
L	For years of schooling of parents		
	intercept	2.794	5.2
S*	school avg. number of years repeated	-2.049	-3.7
L	For learner SES		
	intercept	1.933	0.8
S	principal's teaching load	-0.189	-1.7
E	class methodology value	0.514	1.4
L	For learner's age in years and months		
	intercept	-5.137	-4.9
L	For learner is a girl		
	intercept	9.862	3.2
<u>Analysis of variance</u>		Null model	Above model
		% variance explained	
Level 1 (learner)		3861	3466
Level 2 (school)		3336	1507
Total		7197	4973
			31%

Excluded: years of in-service training; days of in-service training; evaluation intensity; class size squared; teacher SES; level of parent involvement; principal's years of pre-service training; intensity of district support; proximity to urban facilities; degree of teacher latecoming.

Table 3: Summary of effects for historically disadvantaged schools

	Reading models							Mathematics models						
	Values HD			Score gain for HD				Values HD			Score gain for HD			
	10 th	50 th	90 th	OLS all	OLS HD	HLM all	HLM HD	10 th	50 th	90 th	OLS all	OLS HD	HLM all	HLM HD
Variance explained				0.62	0.32	0.64	0.31				0.52	0.17	0.55	0.15
Actual distribution of HD scores	365	439	563					378	440	532				
Predicted mean for HD scores				452	450	451	450				450	448	449	448
E years pre-service training	14	15	15	12	7	8		14	15	15	10		6	
E days of in-service training	0	10	60					0	9	47		2	1*	1
E evaluation intensity	0	0	4	8				0	0	3	5	5		
E class size squared	3600	1764	900		7			3600	1764	900				
E class methodology value	3	6	9	8	16		5	2	4	7	14	10	8	5
E teacher hours in a year	197	709	1209	11*	11	8	7	310	755	1155				
S level of school infrastructure	0.2	2.4	6.6	38	36	38	39				21	16	10	11
S principal's years of pre-service training	14	15	16								3			
S principal's teaching load	19	7	0		7		6				10			
S intensity of district support	1	8	20	3*	4						6*			
S proximity to urban facilities	1	1	3								16	3		
S degree of teacher latecoming	1	1	1	0		0					0	0	0	
L number of years repeated	2	1	0	35	30	31	25				37	34	26	24
L textbooks per learner	0.0	0.5	0.5		3	5	4				5			
L average number of meals per day	1.2	2.9	3.0	4	3	2	4				11	9	2	2
L years of schooling of parents	4	12	17	28	8	27	7				29	8	17	6
L learner SES	0.0	3.3	8.4	21	17	4	13				11	22*	-1	6
L learner's age in years and months	15	13	11	10	10	10	10				24*	20*	4	1*
L learner is a girl	0	1	1	47	36	46	5				28	21		

Excluded: teacher SES; level of parent involvement. Note: The 'score gain' is the predicted gain over the predicted mean when the values for percentiles 1 to 89 are raised (or lowered) to the value of the 90th percentile. * means that the association with performance was the opposite of what was expected, and hence the reverse of the adjustment was run, meaning values from the 11th to the 100th percentile were set to the 10th percentile value. Variance explained refers to the original model in which the coefficients were produced – for instance, the variance explained in the HLM reading model using all the South Africa data was 0.64.

Table 4: Description of variables

E years pre-service training	Average years of schooling/training of (mainly) educators at the school but also of learner's educator.
E days of in-service training	Days of in-service training in last three years (learner's educator).
E evaluation intensity	Intensity of evaluations by the school principal, with once a year evaluations being weighted more. A 0-10 scale used.
E class size squared	Learners in the class squared.
E class methodology value	Value of classroom approaches on 0-10 scale.
E teacher SES	SES of learner's educator on 0-10 scale.
E level of parent involvement	An indicator of the value of parent-educator interactions.
E teacher hours in a year	Number of hours educators teach in a year, without counting absenteeism.
S level of school infrastructure	A score of infrastructure (building and equipment) presence on a 0-10 scale.
S principal's years of pre-service training	Years of schooling/training of the principal.
S principal's teaching load	Hours per week that the principal teaches.
S intensity of district support	Number of visits by 'inspector' to the school in the last 3 years. Access to a nearby resource centre was translated into additional 10 visits.
S proximity to urban facilities	Scale of 1-3, with 3 meaning most urban.
S degree of teacher latecoming	A dummy variable indicating whether the principal believes educator latecoming is a problem.
L number of years repeated	Years that a learner has repeated in the past.
L textbooks per learner	Ratio of textbooks per learner for one subject, with an upper cut-off of 0.5.
L average number of meals per day	Average number of meals eaten per day.
L years of schooling of parents	Years of schooling of mother and father, with mother weighted twice father. Maximum of 7 years added depending on usage of English.
L learner SES	An indicator on a 0-10 scale of learner's SES based on a factor analysis using the presence of four household items.
L learner's age in years and months	The age of the learner in years and in months converted to decimal fractions.
L learner is a girl	A dummy variable with value 1 for girl.

5 CONCLUSIONS

In terms of methodology, two conclusions stand out. Firstly, given the segmented nature of South Africa's schooling system, production functions that are specific to a segment are necessary. Secondly, the HLM tends to understate effects relative to the OLS model.

On the policy front, one should separate effects that imply major budgetary and structural change, and those whose implications are of a more managerial nature. These two categories require very different policy responses. Three managerial issues stand out. Firstly, the allocation of time in schools plays a discernable role. Other things being equal, more teaching time is associated with better test scores. But less teaching time for the principal is also associated with better results, pointing to the importance of time spent managing. Part of the time allocation problem relates to latecoming amongst teachers. Secondly, certain teaching methodologies are clearly associated with better scores, independently of the quantity of training teachers have received. Thirdly, grade repetition approaches, which vary from school to school, are strongly associated with differences in learner performance. Here however, the appropriate policy response needs to be established after further investigation into the repeater issue that goes beyond the typical production function approach. Of course these relationships are reflected in a 2000 dataset. Would they apply in 2006? The answer is arguably yes for the time management and repeater issues, given the absence of vigorous policy interventions in these areas. The teaching methodology issue may be different, given the relatively strong emphasis on the advocacy of better teaching practices (for instance following the 2002 revision of the curriculum).

Three budgetary changes suggested by the models have partly been carried out since 2000. The models point to the importance of infrastructure improvements (especially insofar as this reduces remoteness problems associated with rurality). Annual infrastructure spending on schools increased by 250% in real terms between 2000 and 2006 (National Treasury and Provincial Treasury expenditure reports). The Department of Education's 2005 vision for 'rural schooling' (South Africa, 2005) recognises the specialness of rural schools. The need for more adequate textbook provisioning would to a large degree have been addressed by the 210% real increase in spending (2000 to 2006) on non-personnel recurrent items. However, with respect to better learner nutrition, budgets have remained more or less static in real terms during the six year period, with per learner spending per school day remaining at about 68 cents (in 2006 rand terms and assuming coverage of 50% of all learners, or around 5.5 million learners). However, the true test will lie not in budgetary allocations but in improvements in learner performance scores in future runs of SACMEQ and other, similar standardised testing programmes.

REFERENCES

- Bhorat, H. & Oosthuizen, M. (2006). *Determinants of Grade 12 pass rates in the post-apartheid South African schooling system*. Dakar: Sisera. Available from: <http://www.idrc.ca/uploads/user-S/114787836912006_6_DPRU_SAGA_Bhorat_Determinant.pdf> [Accessed October 2006].
- Case, A. & Deaton, A. (1999). School inputs and educational outcomes in South Africa. *Quarterly Journal of Economics*, 114(3): 1047-1084.
- Crouch, L. & Mabogoane, T. (1998). When the residuals matter more than the coefficients: An educational perspective. *Journal of Studies in Economics and Econometrics*, 22(2).
- Crouch, L. & Perry, H. (2002) *Determinants of learning and learning change in DDSP schools' Grade 3 numeracy and literacy assessment results*. District Development Support Programme, Pretoria. Unpublished mimeograph.
- Eisemon, T.O. (1997). *Reducing repetition: Issues and strategies*. Paris: IIEP. Available from: <<http://www.unesco.org>> [Accessed December 2006].
- Glewwe, P., Kremer, M. & Moulin, S. (2000). Textbooks and test scores: Evidence from a prospective evaluation in Kenya. Washington, World Bank. Available from: <http://post.economics.harvard.edu/faculty/kremer/webpapers/Textbooks_Test_Scores.pdf> [Accessed 5 June 2004].
- Gustafsson, M. (2005). *The relationships between schooling inputs and outputs in South Africa: Methodologies and policy recommendations based on the 2000 SACMEQ dataset*. Harare: SACMEQ. Available from: <www.sacmeq.org> [Accessed October 2005].
- Gustafsson, M. (2006). *School production modelling to strengthen government monitoring programmes in developing countries*. Masters thesis. University of Stellenbosch. Available from: <<http://ir.sun.ac.za/dspace/items-by-author?author=Gustafsson%2C+Martin+Anders>> [Accessed December 2006].
- Gustafsson, M. & Patel, F. (2006). Undoing the apartheid legacy: Pro-poor spending shifts in the South African public school system. *Perspectives in Education*, 24(2): 65-77.

- Hanushek, E.A. (2002). *The failure of input-based schooling policies*. Washington: National Bureau of Economic Research. Available from: <<http://papers.nber.org/papers/w9040.pdf>> [Accessed November 2005].
- Hanushek, E.A. & Luque, J.A. (2003). Efficiency and equity in schools around the world. *Economics of Education Review*, 22: 481-502.
- Johnes, J. (2004). Efficiency measurement. In Geraint Johnes and Jill Johnes (ed.), *International handbook on the economics of education*. Cheltenham: Edward Elgar: 613-742.
- Lee, J. & Barro, R.J. (2000). *Schooling quality in a cross-section of countries*. Seoul: Korea University. Available from: <[http://econ.korea.ac.kr/prof/jwlee/files/school%20quality\(1\).pdf](http://econ.korea.ac.kr/prof/jwlee/files/school%20quality(1).pdf)> [Accessed July 2005].
- Monk, D.H. (1990). *Educational finance: An economic approach*. New York: McGraw-Hill.
- Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks: Sage.
- South Africa: Department of Education (2004). *Strategic plan 2004-2006*. Pretoria.
- South Africa: Department of Education. (2005). *Report of the Ministerial Committee on Rural Education: A new vision for rural schooling*. Pretoria. Available from: <<http://www.education.gov.za/content/documents/677.pdf>> [Accessed 15 June 2005].
- UNESCO (2005). *Education for All global monitoring report 2005: The quality imperative*. Paris.
- Van der Berg, S. (2005). *How effective are poor schools? Poverty and educational outcomes in South Africa*. Stellenbosch: University of Stellenbosch. Available from: <<http://academic.sun.ac.za/econ/images/SvdBerg-SACMEQ-paper.doc>> [Accessed March 2006].
- Van der Berg, S. & Burger, R. (2003). *Education and socio-economic differentials: A study of school performance in the Western Cape*. University of Stellenbosch.
- Willms, J.D. & Somers, M. (2001). Family, classroom and school effects on children's educational outcomes in Latin America. *School effectiveness and school improvement*, 12(4): 401-445.