RESOURCES
and outcomes in public schools

The case of South Africa

Russell Wildeman

Resources and Outcomes in Public Schools

The case of South Africa

By Russell Wildeman

idasa
AN AFRICAN DEMOCRACY INSTITUTE

2010

Idasa would like to thank the Department for International Development (DFID) for the financial assistance in this project.

DFID Department for International Development

ISBN 978-1-920409-17-3

First published 2010

Editing by Paul Wise

Design and cover design by Welma Odendaal

Cover photograph: © Peeter Viisimaa/iStock Photo

Production by media@idasa

Bound and printed by Top Copy, Cape Town

Contents

Abbreviations and acronyms

LEA	local education authority (UK)
MAR	missingness at random
OLS	ordinary least squares
PPE	per pupil expenditure
SACMER	Southern and Eastern Africa Consortium for Monitoring Education
SES	socio-economic status

Foreword

Education research has occupied a special place in the research and advocacy work that Idasa has conducted over the last ten years.

This paper continues this work and is the first in a series that looks at issues that affect the performance of public schools in South Africa. Although most of the earlier Idasa research has employed a qualitative research methodology, this and future research will have an explicitly quantitative dimension and will draw upon local, regional and international surveys.

Regionally, this research will be extended to six African countries, Ghana, Kenya, Malawi, Swaziland, Uganda and Zambia. Idasa has implementing partners in each of the target countries and the Department for International Development (DFID) funds a five-year Right to Know, Right to Education Project in the six countries. This five-year project aims to establish an explicit right of access to information dimension in education policies, increase national stakeholders familiarity with applied budget analysis and advocate for change in the governance and funding of public primary schools.

For the South African leg of the research, the following key areas will be examined:

- The role of socio-economic segregation in determining academic outcomes in public schools;
- Social capital and school improvement and how relevant this is for poor schools in South Africa and the region;
- The notion of effectiveness in public schools and whether we have the data to make comprehensive performance judgments of public schools; and
- The possibilities for the establishment of a truly national longitudinal study of education and education change and detailing its importance to relevant policy debates in South Africa and the region.

In the writing of this paper, I was generously supported by my colleagues in the Economic Governance Programme. Wellington Jogo provided technical assistance while Carmen Alpin and Manyewu Mutamba read earlier drafts of the paper. Douglas Racionzer also provided valuable input on the target audience for this paper and suggested innovative ways of making the paper more accessible. One of the immediate spin-offs of the paper is that the Economic Governance Programme has now established a Data Analysis Group that will routinely meet to discuss technical and statistical issues related to the models we want to estimate. This is an exciting development and should add to the public discourse on education and other relevant social services delivery areas in South Africa and beyond.

Russell Wildeman

Programme Manager

Economic Governance Programme

Idasa

2010

Executive summary

The objective of this paper is to examine the relationship between school resources and outcomes. While it does not comprehensively model the determinants of learner performance, other than school resources, it does provide insights into the factors that affect learner performance for Grade 6 learners. This paper used data from the Department of Education's Systemic Evaluation Survey of Grade 6 learners in South African public primary schools. The survey results and the data were published in 2004. The key findings of the present paper are:

- The index of school resources is positively related to outcomes in language, mathematics and science. Although the magnitude of the resource effects is relatively small, a case can be made that the results are educationally and practically significant.
- Class size as reported by the school principals of the primary schools that participated in the survey is a significant predictor of academic outcomes in all three subject areas, but its effects are much smaller compared with the index of school resources.
- The data do not offer evidence that the effects of class sizes are different for urban primary schools in language and science performance. However, the negative effect of class size on mathematics performance appears larger for urban primary schools.
- The negative effect of class size on the three outcomes seems to be substantially muted in high socio-economic status (SES) schools. This differential effect is large enough to reverse the direction of the relationship between class size and outcomes in language, mathematics and science for high-SES schools.
- There are large (unadjusted) mean differences in performance between rural and urban public primary schools. However, once the average SES of primary schools is controlled for, the large mean differences are substantially reduced for language and science outcomes. However, urban schools are still predicted to do significantly better than their rural counterparts in mathematics.
- The average SES of primary schools has a huge impact on the overall performance of learners. A one-unit increase in the average SES of primary schools is predicted to boost individual learner performance by almost two points in language and just short of one full point in science.
- However, the individual SES of students does not appear to affect learner performance significantly. This is a counter-intuitive finding and may hinge on the quality of measures used in the sample data.
- Primary schools that have a larger intake of learners from different socio-economic backgrounds are predicted to do worse than schools that have more homogeneous learner intakes. This may reflect the strong role of between-school socio-economic segregation in the primary schooling system and the deliberate measures that many middle-class schools take to exclude learners whose parents cannot afford above-average school fees.
- Girl learners are predicted to do substantially better than boy learners in language and science, but gender differences are statistically insignificant for performance in mathematics. The gender achievement gap is particularly large for language, where girl learners are predicted to have a three-point advantage over boy learners.
- Older learners are predicted to do significantly worse than their appropriately aged cohorts, and the achievement differences range from a penalty of half a point for mathematics to almost two points for language.

Attendance at any pre-primary educational institution is associated with positive gains across the three outcomes. The fact that this variable measures only access to pre-primary education rather than the quality of pre-primary education makes this finding even more interesting and relevant.

Internationally, the debate about the importance of resources in determining outcomes has produced mixed results. The fact that researchers use different methodologies and apply different research measurement standards partly explains the inconsistent results. However, the consensus that has formed points to the importance of longitudinal data, rigorous controls at the pupil and school levels (especially prior academic ability) and data at different levels of the schooling system. Some researchers go so far as to suggest that school resources measured should be those that reach children instead of budgeted resources that may never reach children because of underspending or fraud. Similar arguments have been made about the variability of "class size", which is often measured as if it were a constant for learners in a school, while we know that in practice class size changes regularly during a normal school day as learners – particularly in secondary schools – are tracked into different subject and academic streams.

Researchers have also begun to make the case that simple correlations between school resources and outcomes are not rigorous because the resource allocation process and the outcome processes are potentially intertwined, leading to spurious negative or statistically insignificant results. A case in point is class size, where schools often have "rules" for allocating children to various-sized classes based on informal and formal assessments. If lesser-ability children are put into smaller classes and higher-ability children into larger classes because of the latter's ability to derive more value from instruction, a simple correlation may suggest a positive relationship between class size and outcomes. However, this ignores the class allocation rules of schools; and where appropriate controls such as data on prior academic ability are available, researchers have found substantial negative relationships between class size and academic outcomes. Similar arguments have been advanced for school resources defined more broadly, and hence it becomes imperative to model such processes explicitly before definitive conclusions are made about the resources and outcomes nexus.

South African quantitative education research literature is very small, lacks diversity and has not grappled with some of the methodological challenges in estimating resources and outcome relationships.

Research in South Africa on the relationship between resources and outcomes has been largely dismissive of the importance of school resources. There are notable exceptions where resources more generally and class size more specifically have been found to be significant predictors of learner performance.

The South African quantitative education research literature is very small, lacks diversity and has not grappled with some of the methodological challenges in estimating resources and outcome relationships. As a result, the same finding about the relative unimportance of resources has been made repeatedly, thus reinforcing the resolve of those who believe that South African public schools are inefficient users of scarce state resources. While very few would argue that wastefulness and inefficiency do not exist, this paper also establishes that we do not have the data to make strong statements about schools' effectiveness or their use of school resources. At the very least we should be comparing learner intakes across schools, but no South African education survey that this writer is aware of has this data, which excludes from consideration one of the most potent variables explaining learner performance and, by ex-

Resources are important, but their availability and efficient use will not by themselves bridge the education quality gap in South Africa.

tension, mean academic performance. When such measures have been available in international studies, large school effects (or differences) that were evident before adjusting for prior academic ability often disappear.

The most obvious limitation of the present paper (stemming in part from the quality of the data used) is that the potential endogeneity of resources with respect to the three outcomes has not been taken into account. Second, the data are cross-sectional and therefore questions must persist about whether the same relationships endure over time. Third, the absence of data on prior academic achievement makes it impossible to say anything meaningful about school effectiveness or schools' differential use of resources. It is vital that we compare the same quality intakes over schools and then draw "value-added" conclusions about how schools retard or accelerate learner progress. Fourth, the large between-school socio-economic segregation situation has not been analysed in this paper, and separate research is needed to consider the resources and performance implications of our divided public primary and secondary schooling system.

Resources are important, but their availability and efficient use will not by themselves bridge the education quality gap in South Africa. Two important factors counter an approach based exclusively on resources, namely the high between-school socio-economic segregation and the fact that we know so little about what happens in primary school classrooms. Better resource conditions will undoubtedly help, but we now need to shine the investigative torch into the classroom, which is, after all, the immediate point of contact for learning and development. The fact that the present data survey does not allow one to match teachers to classes and learners is symptomatic of an overall avoidance syndrome regarding classroom teaching. This problem involves all stakeholders and must be addressed at all levels of the system. Without such actions, resource benefits will continue to accrue to very specific groups of learners and schools and the desired outcomes will remain glaringly unequal.

Introduction

After more than a decade and a half of reforms to the funding of South African public schools, concerns persist about the efficient use of resources in schools that do not perform well on standardised tests of achievement. There seems to be near universal consensus that further funding increases to public schools that serve poor communities will not help arrest performance deficits in these schools. South Africa's persistently poor performance in international achievement tests provides proponents of this view with powerful ammunition, especially when performance in South Africa's public schools is compared with that of its poorer neighbours in Southern and East Africa.

The government's response to this line of research has been revealing. On the one hand, it appears to have accepted the view that a large percentage of South African public schools continues to let poor learners and poor communities down. This view has found strong favour, especially with the National Treasury under the former Minister of Finance, Trevor Manuel. Even the national Department of Education has tried several strategies to get poorly performing schools to work. On the other hand, the same government has continued to pursue a resources agenda that channels an increasingly large share of education resources to poor and poorly performing schools.

Thus, on the surface at least, one may argue that the government has not been completely swayed by those who believe that better resources will not make a difference to schools' overall academic achievement. This default position could also be due to political pressures in the ruling African National Congress and its tripartite alliance with the Congress of South African Trade Unions and the South African Communist Party, which induce education authorities, to pursue a political and policy agenda over which they do not have control.

Examining the present policy landscape, it seems clear that quantitative researchers that question the resources-outcomes nexus have not succeeded in deflecting the South African government from its stated aim of providing better resources to poor and poorly performing schools.

I would argue that although the present funding system is far from perfect, we should continue to advocate the better funding of public schools, especially while South African quantitative education research is still in its formative stages regarding access to high-quality data surveys, estimation methods and research sophistication. I will have more to say about this in my short review of the international and local literature, but the critical point is that we do not yet have a critical mass of quantitative education researchers who carefully review and study one another's work or who are steeped in the recent methodological trends and innovations so evident in the international literature.

I am also not aware of any major public debates within the small community of researchers who pursue quantitative research in education, and this is symptomatic of the challenges facing researchers in this area

Quantitative researchers who question the resources-outcomes nexus have not succeeded in deflecting the government from its stated aim of providing better resources to poor and poorly performing schools.

(resources-outcomes) and other relevant areas. Such debates tend to sharpen the way researchers measure variables of interest, drive the demand for better-quality surveys and help researchers construct careful policy advice for policy-makers who do not understand the intricacies of high-end research. Also, the gap between researchers in education employing qualitative and quantitative methods has not helped the development of sophisticated quantitative research that could, as Harvey Goldstein (1981) argues, better approximate the complex realities that define education systems.

This paper pursues the following main aims:

- It sets out to provide a brief review of the international and local literature that examines the relationship between resources and outcomes (section 2). None of the trenchant critiques of quantitative methodologies are included in this review and recent debates about how to define the "effectiveness" of schools are only mentioned briefly. There is a much larger literature on this subject, but for present purposes this short review attempts to highlight key methodological and empirical findings in this area.
- It uses South African Grade 6 data from the Systemic Evaluation Survey released in 2004 and offers an exploratory analysis of the relationships between resources and outcomes in South African public primary schools (section 3).
- It suggests areas for further research that would not only address the controversial issue of school resources and outcomes but also begin to deal with the large socio-economic gaps that define public primary and secondary schooling in South Africa.

Review of international and local literature on the relationship between school resources and academic outcomes[1]

Levacic and Vignoles (2002) note that quantitative education modelling that examines the relationships between inputs and outputs can be traced back to the seminal Coleman Report released in 1996 in the United States of America. In their view, this report has spawned two traditions of quantitative education research, namely school effectiveness research and the use of education production functions. According to the authors, school effectiveness research dominated by educationalists has largely been preoccupied with measuring the effect of different schools on student attainment and searching for school process factors associated with differential effectiveness (Levacic and Vignoles, 2002: 315).

The other research tradition, education production functions, has largely focused on testing hypotheses about the "causal" relationship between school resources and school outcomes. Levacic and Vignoles (2002) argue that in the latter approach, factors other than resource inputs associated with student achievement are often treated as confounding factors, although this approach acknowledges the importance of individual student characteristics.

Education production function studies have generally paid little attention to internal school processes and typically treat the school as a black box. Levacic and Vignoles (2002) state that the model of school effectiveness research is now widely accepted in its conceptualisation of the research process as focusing on context, inputs, school processes and individual and school outcomes.

In his seminal study of the relationship between school inputs and outcomes, Hanushek (1986: 1142) argues that education production function studies are seldom designed to deal with the complex policy issues that are central to the investigation of schooling. Murnane (1991) is more emphatic in his rejection of education production functions as a viable method of answering questions about the relationship between school resources and outcomes. His main concern is about the problems of causality.

In the United States and other First World countries, poor learners and schools receive compensatory funding. If one accepts that the resource allocation process and characteristics of such schooling communities are related to performance, then the estimated relationship between school resources and outcomes is likely to be a spurious negative relationship. Of course, the same logic applies to public schools that receive less government funding, but attract higher-quality student intakes annually. Case and Deaton (1999) argue that positive relationships between school resources and outcomes in affluent communities may have more to do with unobserved parental tastes for education, and it is hard to disentangle the effects of such tastes from those of school inputs.

Murnane (1991) makes the important point that education production function research has often been used to delegitimise progressive resource agendas by showing that schooling inputs bear no relationship to relevant academic outcomes. However, he cites examples of resource variables that are rewarded in private schooling and business contexts (such as loyalty and years of teaching experience) for which research has

Increased spending by itself will not lead to better academic performance. According to Murnane, the behaviour of key role players (teachers, managers, district officials) as well as governance and hiring practices must be changed to maximise the effect of targeted resource increases to public schools.

found a weak direct connection with outcomes. Because private institutions operate and survive under conditions of competitive pressures, these decisions are less scrutinised and criticised. This begs the question: why should the relevance and usefulness of the same resource inputs in public schools be questioned? Murnane does moderate his arguments about spending in public schools by insisting that increased spending by itself will not lead to better academic performance. According to him, the behaviour of key role players (teachers, managers, district officials) as well as governance and hiring practices must be changed to maximise the effect of targeted resource increases to public schools.

So far I have focused on some of the methodological issues plaguing quantitative education research that probes the school resources and outcomes nexus, but questions remain about the kind and quality of the data that should be used to draw credible conclusions about the relationships between school resources and outcomes. In their review of education resource studies in the United Kingdom, Levacic and Vignoles (2002) conclude that such studies must have the following features:

- They should include data at the pupil level on measured academic achievement with controls for prior academic achievement (proxy for ability).
- They should also include individual and family data on pupils' background and school resources variables at least, measured at the school level.
- Endogeneity problems resulting from the correlation between school resources and the residuals of the dependent (criterion) variable must be resolved through the use of instrumental variables where available.
- Researchers must rigorously test the functional form of the relationship.
- Multilevel modelling or equivalent econometric models should be used to take the "nested" structure of school data into account.

Greenwald, Hedges and Laine (1996: 385) go one step further in their remarks about school resources by insisting that researchers measure the school resources that reach specific children rather than using the school or district averages. One can add to this list the availability of longitudinal data because quantitative education researchers are deeply suspicious of the reliability and durability of cross-sectional survey results and analyses. From this perspective, relationships that hold robustly over time and where appropriate statistical controls are in place offer far more persuasive evidence.

Although it is true to say that suspicion about the value of schools generally and school resources more specifically in determining outcomes probably originated with the Coleman Report, the influential review by Hanushek (1986) stands as a landmark in this controversial debate.

Hanushek reviewed 38 articles and books, including 187 regression equations, which formed the basis for his sample and his pessimistic conclusions about the relationships between school resource inputs and outcomes. He poses what he considers the "puzzle" in the following way: in spite of the constantly rising costs and quality of the inputs of public schools in the United States, these trends have not been matched by a corresponding improvement in the performance of students. This leads him to conclude that the resources and outcomes relationship is at best ambiguous and at worst negative.

His method was to organise the relationship (coefficients) between school resources and school outcomes in his sample into five categories: statistically significant coefficients, statistically insignificant coefficients, positive or negative signs on the coefficients, and statistically insignificant coefficients for which the direction of the coefficients could not be determined from the published books and articles he had examined. His conclusions were based on a process of adding the various categories (x percent of a resource variable was found statistically significant positively or negatively, etc). Because such a low proportion of the resource variables demonstrated a positive statistically significant relationship with outcomes and some showed negative statistically significant relationships, Hanushek formulated the following finding, which was cited subsequently by researchers and government officials opposed to the better resourcing of public schools:

> Without systematic tabulation of the results of the various studies, it would be easy to conclude that the findings of the studies are inconsistent. But there is a consistency to the results: *There appears to be no strong or systematic relationship between school expenditures and student performance.* This is the case when expenditures are decomposed into underlying determinants and when expenditures are considered in the aggregate (Hanuschek, 1986: 1162 – italics in original).

Although Murnane (1991) does not discuss the seminal Hanushek (1986) article in detail, it is clear that most of his comments and critique apply to it. However, the first systematic response to the article was provided by Hedges, Laine and Greenwald (1994b). This occasioned a debate between Hedges and his co-writers and Hanushek (chronicled in the education academic journal *Educational Researcher*), and follow-up research was subsequently done by Greenwald, Hedges and Laine (1996) in which the same authors selected their own sample of studies and conducted meta-analyses[2] of the data. Hedges and his colleagues attacked the review by Hanushek on three grounds:

- Hanushek's vote-counting (his tallying method) is suspect because it cannot correct for weak statistical power in studies that form part of his universe of studies. Therefore, if statistically significant positive relations are present in the population from which the sample is drawn, but the sample has low statistical power, reviewers could be forced to draw the incorrect conclusion that there are no such relations. This increases the probability of committing a type II error.
- If one accepts the hypothesis of no relationship between resource variables and school outcomes, then one would ordinarily expect to have, by chance, statistically significant results 5% of the time (equally divided between positive and negative relationships). However, in Hanushek's own results, the percentage of studies having positive coefficients ranged from 2.3 to 7 times the 5% that would be expected through chance alone if school resource variables and outcomes were completely unrelated.
- The vote-counting methodology that Hanushek employs does not allow for the calculation of effect sizes, which raises the question of the distinction between practical educational significance and statistical significance.

Hedges, Laine and Greenwald employed a meta-analytic technique, which involved combinations of significance tests and the calculation of effect sizes. The first combined the p-values from the studies that used different research designs and used a statistical test[3] to determine whether the data were consistent with the null hypothesis of no relations in all the studies. For the calculation of the effect sizes for each of the resource variables, the authors used the partial standardised regression coefficients, because the variables were not measured on the same scale. In their reanalysis, the authors found strong support for some positive effects of school resource inputs and little support for the existence of negative effects (Hedges, Laine and Greenwald, 1994a: 13). Also, in their effect size analyses, they found evidence of the practical importance of especially the per pupil expenditure (PPE) and teacher experience variables. Finally, the median effect sizes for the resource variables were found to be positive for most school resources variables, but the teacher education variable had a negative sign.

Greenwald, Hedges and Laine (1996) constructed a new universe of education production functions from various academic publications (including some of the Hanushek studies) and replicated the initial findings in their original paper (Hedges, Laine and Greenwald, 1994a). The 1996 paper concludes that school resources are systematically related to student achievement and relations are large enough to be educationally important. More importantly, the authors draw the speculative conclusion that school resources act as a substitute for declining levels of social capital, because their analyses noted positive performance changes for minority students in the US.

The debate between Hanushek and researchers using meta-analysis has begun – at least in the US – to dent the consensus about the unimportance of school resources in determining educational outcomes. Equally interesting is that researchers examining the resources-outcomes nexus in the wake of this debate have not chosen to use research synthesis techniques, but have confined themselves to doing single-study analyses from available survey data.

Goldhaber and Brewer (1997) use a series of regression models (both fixed and single-level random models) to model the relationship between school resource variables and academic outcomes. Their main focus is on explaining why pertinent variables such as teacher characteristics (training, education and specialisation) and school resources more broadly do not appear to be statistically positive in a large number of education production function studies. Their findings suggest that students of more experienced teachers have higher test scores; however, the crucial element appears not to be teachers' qualifications generally, but qualifications specifically in mathematics (the subject of their analysis). In their research, having a master's degree without the right certification in mathematics did not show positively significant results.

Also, they found that observable teacher characteristics often noted in large-scale surveys (such as sex, race, gender, degree level and experience) explain a small proportion of the variation in outcomes, but un-observable teacher and school variables do appear to play a key role in accounting for performance differentials. They came to this conclusion after they fitted fixed teacher effects in their models, which raised the coefficient of determination by a significant margin. Their paper mentions auxiliary regressions on which the fixed teacher effect was regressed using observable teacher characteristics. The only basis for their touting of the role of un-observable teacher attributes can be that observable teacher characteristics do not explain all the variations in the fixed teacher effect. So while they do not know what these characteristics are, their work supports the idea that school resources, broadly defined, do have an impact on student performance.

Blatchford et al. (2002) tackle the contentious issue of class size (a school resource variable) and its relationship to attainment for children in the first three years of their schooling. A key feature of this study is that it is longitudinal and controls for the entry ability of young children. The latter feature is especially important in controlling partly for the non-random allocation to classes and class sizes of children whose "abilities" are determined formally and informally. This is vital because if research ignores the class allocation process, it may find correlations

Blatchford et al. noted strong relationships between class size and attainment in literacy and mathematics, with the overall pattern being decreasing performance as the class sizes increased.

between class size and attainment that are, in fact, partly or wholly explained by the initial class allocation process.

Blatchford et al. noted strong relationships between class size and attainment in literacy and mathematics, with the overall pattern being decreasing performance as the class sizes increased. However, they found a threshold class size limit beyond which class size did not appear to matter. This suggests that the effect of a given reduction in class size depends on the actual class size itself. According to the authors, the optimum class size appears to differ across research in the USA and the UK. This study also found that small class sizes benefit children who are most in need academically and those with lower school entry scores.

Levacic et al. (2005) examined the relationship between school resources and academic outcomes in mathematics, science and English in Key Stage 3 in English schools. The authors adopted an instrumental variable approach to counter the effects of endogenous resource variables. In essence, they regressed the resource variables (PPE and pupil-teacher ratios) on two instruments, namely the political party in control of the local education authority (LEA) and lagged school size. These variables were hypothesised to have an impact on overall resourcing levels, but one that was unrelated to the criterion variables (academic achievement in mathematics, science and English).

In both their ordinary least squares (OLS) estimation and multilevel instrumental variable specification, the authors found that PPE had a statistically significant effect on Key Stage 3 attainment in mathematics and science. They had similar findings for the effect of pupil-teacher ratios on mathematics and science attainment. However, the effects on attainment using the instrumental variable approach were ten times larger than those using the OLS approach. The interesting and surprising finding is that none of the resource variables had a statistically positive effect on attainment in English.

Levacic et al. also produced evidence that resource effects were larger for students of poor families with above-average ability[4], and smaller resource effects were found for students with average academic abilities in mathematics and science. In spite of the significance of the resource effects, the authors note that those effects were relatively small and suggest that targeted resource interventions were more likely to have an effect on performance than a large general increase in resources.

Steele, Vignoles and Jenkins (2007) extended the model estimated by Levacic et al. (2005) and estimated a full multilevel, multivariate simultaneous equation model to account for the nested structure of the data and the endogenous nature of the resource variables. They used the same instrumental variable approach as that of Levacic et al. to counter the endogenous nature of the resource variables, but applied an adapted version of the multivariate multilevel model, which estimates two or more criterion (dependent) variables at the same time.

The authors also noted that their research differed from the traditional instrumental variable approach in the assumptions made about the level at which the selection effects operated. In the standard instrumental variable approach, which estimates equations for the dependent variables (outcomes) of interest, the equations are linked via correlated residuals defined at the lowest level of observation, namely the pupil level. But Steele, Vignoles and Jenkins noted that this approach was incorrect in the present case because it erroneously treated school resources as a pupil-level variable and did not recognise that endogeneity arises due to unobservables at the school or local educating authority (LEA) level rather than at the pupil level. In their analyses, the authors found evidence of the endogeneity of school resources with respect to mathematics and science, but not for English.

Furthermore, the authors found a positive statistically significant effect of resources (PPE and pupil-teacher ratio) on attainment in mathematics and science. As in the study by Levacic et al. (2005), these effects were found to be much larger once endogeneity had been accounted for. However, for attainment in English, the results indicated a negative effect of PPE on English attainment and no significant effect of the pupil-teacher ratio on the same subject.

Steele, Vignoles and Jenkins conclude that policies aimed at reducing the pupil-teacher ratio may be effective, especially for mathematics and science. Also, in spite of the now famous conclusion by Hanushek (1986) that public schools do not use resources efficiently, their results show the opposite, notwithstanding their qualification that the resource effects are still modest.

The sophisticated modelling by Steele, Vignoles and Jenkins of the resources-outcome nexus is a far cry from the black-box approach often adopted in quantitative education research, but their results do not warrant a wholesale increase in education expenditures. Most of the research I have reviewed suggests targeted interventions, which imply a continuation of funding policies aimed at helping those who are disadvantaged in our societies. However, this does not mean that a basic standard for education funding is impossible (a so-called output adequacy or input standard), but simply that even if we were to adopt such a standard, differentiation may still be required to mitigate the impact of unequal societies.

Raudenbush (2009), in a lecture dedicated to reviewing the effects of legal desegregation and affirmative action in public schools and higher education institutions in the United States, notes that school quality for African Americans followed in the wake of broad societal changes. This raises the question whether school improvement by itself can contribute to a reduction in racial and socio-economic inequalities. He answers this question in the affirmative, arguing that learners from disadvantaged backgrounds need more and better-quality schooling.

Raudenbush reviews studies showing that early childhood development and simply being in school matter most for poor learners whose parents have low levels of formal education. On the quality side, he says that three conventional school resources, namely class size, teacher experience and teacher knowledge, have been found to have positive and statistically significant associations with academic outcomes across a broad range of studies.

However, like Murnane (1991), he believes that the potential contributions of these resource variables will be reduced if teaching practices are not organised differently. In this regard, he contrasts what he calls a privatised idiosyncratic instructional regime with a shared systematic instructional practice. The former follows the traditional model of teacher autonomy, where there is little scrutiny and accountability of teachers, whereas the latter model institutionalises scrutiny and shared learning. For Raudenbush, the critical contact point in determining the usefulness of school resources changes is the classroom, and he advocates a reorganisation of classroom teaching practices that will maximise such resource benefits.

South African quantitative education research has been carried out by a small group of mostly university-based researchers in South Africa and abroad. Because my primary focus is on the resources-outcome nexus, a great deal of the local writing will not be reviewed, and obvious space considerations prevent me from selecting more than a few of the studies.

Case and Deaton (1999) have examined the relationship between class size and a range of academic outcomes including school attendance, educational attainment and achievement on cognitive standardised tests. Being aware of the possible endogeneity of the class size variable, Case and Deaton make the argument that black families have had little control over resource allocation decisions in public schools. Furthermore, they are able to demonstrate in the sample data no statistically significant relationship between household expenditures and the educational levels of the household head in black families on one hand, and the pupil-teacher ratio (proxy for class size) on the other [PC B3]. The authors need this evidence to counter alternative explanations of the relationship between class size and education outcomes, should a statistically significant relationship emerge for black students.

In their findings, class size does not matter for white students but has a large negative effect on educational outcome measures for black students. They also argue that class size appears to have an effect on the age at which young children start school (intercept) and the rate at which they progress through schools (slope). While their sample for educational attendance and attainment is sufficiently large, the corresponding sample used in estimating the relationship between class size and achievement in mathematics and numeracy is very small, which should encourage caution about generalisation of the latter results.

Crouch and Mabogoane (1998b) criticise the Case and Deaton class size findings, arguing that the explanatory power of their regression models is relatively low and that they measure school resources at the magisterial level. In my view, the first criticism is less damaging because Case and Deaton set out to test a relatively small number of (available) explanatory factors, and the coefficient of determination is not the only criterion by which one judges empirical results. However, the fact that resources are measured at the magisterial level is more serious and clearly does not conform with today's research measurement standards.

An interesting feature of the Case and Deaton study is that the intended audience for this article appears to be international (possibly North American) as is evidenced by the quality of the academic journal (*Quarterly Journal of Economics*) in which it was published. Also, their contribution can be seen as part of a long list of replies to the seminal Hanushek (1986) article as well as an indication of the growing importance of meta-analysis as a viable research and analysis technique.

Crouch and Mabogoane (1998a) combined four databases to estimate the relationship between poverty and school factors on the grade 12 pass rates for the Gauteng province. The authors were keen to point out that too much emphasis had been accorded to the redistribution of resources in the post-apartheid period instead of the "wise" management of school resources. The final aggregated regression model estimated shows large effects for poverty and place as well as for the average qualifications of teachers.

Most of the traditional school resources variables used (school buildings, class size, teacher experience) were not found to be statistically significant, and only the computer variable showed a small statistically significant effect. The results led the authors to claim that the disproportionate attention given to the redistribution of numerical resources was misplaced and policy-makers were urged to focus on quality and management factors. However, the authors were transparent in pointing to the limitations of the data, even though this did not stop them from making this strong claim:

> In spite of these [data and methodological] problems the results are strong

> enough that their implications cannot be ignored – they pass the economists' inter-ocular test: they hit one right between the eyes (Crouch and Mabogoane, 1998a: 3).

A further interesting feature of this paper is the authors' use of comparisons between the predicted and the actual performance of schools. Using this methodology, they were able to identify "over-performers" relative to what their regression models predicted for such schools. This led them to speculate as to the kinds of resources support and interventions that such schools would need.

In my view, however, much depends on how rigorously their model is specified and whether they have the data to make such strong recommendations. This approach is conceptually similar to taking residuals from aggregated school regression models and using the results as an indicator of efficiency. I will have more to say about this methodology when reviewing a paper by the same authors that explicitly examines residuals from school regression models. For now, it suffices to refer to recent research using multivariate multilevel regression (see Yang et al., 2002), where the authors make the important point that school effectiveness cannot be established by using one measure of academic outcome only.

Yang et al. were able to show moderate to weak residual correlations in the performance of schools using different mathematics outcomes at the A and AS levels. Schools could be effective in mathematics, but might not do as well in language and science, for example. The fact that the Crouch and Mabogoane paper used only overall grade 12 pass rates renders their judgment of the overall effectiveness of schools problematic. We need to know whether a high overall pass rate is correlated with high performance in all the key subjects that are offered at the grade 12 level, and such evidence is not available from the Crouch and Mabogoane (1998a) paper.

Crouch and Mabogoane (1998b) analysed a sample of 303 secondary schools and developed an absolute achievement index to rank schools on a number of weighted variables. These variables were mostly passes in mathematics and English, an overall grade 12 pass, and sport and cultural variables. Each of these variables was then regressed on two independent variables, namely school resources (class size, number of head of departments, etc.) and a measure of advantage that included area of residence and the former education department to which these schools belonged. The residuals from each of these mini-regression models were stored and summed using the same weighting scheme that was applied in the construction of the absolute achievement index. These residuals were used to create a relative achievement index. This process enabled the authors to compare performance on both the absolute and the relative achievement indices to describe situations where schools do much better than expected, given their resources and levels of social and economic advantage.

Crouch and Mabogoane argue that their analyses show that only 25% of the variability in the outcomes is explained by resources availability and that caution should apply in assessing the contribution of resources to addressing unequal educational outcomes. The authors urge further reviews and analyses and also call for qualitative analyses in which the performance of "outlier" schools is benchmarked carefully (1998b: 15).

What is admirable about this paper is the authors' acknowledgement of the limitations of the data and the data-generating process (self-selection problems, creaming off of high-ability students etc.) and their appeal for better-quality data to allow more rigorous testing of the data.

The use of residuals from aggregated regression models as an indicator of the efficiency of schools has come under attack, especially from researchers in the UK and the USA. Woodhouse and Goldstein (1988) argue that the use of residuals from such models has no basis in theory and in practice. They base their arguments on reanalyses of such models in the UK contexts and come to the following conclusions:

- Aggregate models assume that the within-school and between-regression relationships are the same, thus opening such models to charges of aggregation bias.[5]
- As model fit improves (measured by the coefficient of determination), residual rankings of schools change markedly. Thus it is impossible to discern or make decisions about "best fit" under such circumstances, and the divergence in the residual ranking of schools causes serious problems of interpretation.
- The relative position of the school residual is affected by minor changes made to variables in the aggregated regression model, whether it is the dependent or the independent variables that are transformed using non-linear transformations.
- The rank order of the school residuals also depends on whether they have been standardised, because residuals will be differently distributed with different variances at different points in the data space (Woodhouse and Goldstein, 1988: 302).

In the Crouch and Mabogoane (1998b) paper, the authors do not provide evidence of alternative regression specifications and thereby implicitly assume that this is the best model, which is problematic. There may well be other models that could be fitted to the data with the same (or a better) fit, yet produce different residual rankings. Also, the lack of data at the individual level, especially prior academic achievement data (proxy for ability), makes the use of school residuals as efficiency indicators even more problematic, given the powerful predictive power of this variable. In the authors' defence, however, they make a strong case at the end of their paper for the establishment of high-quality survey data to answer the hard policy questions that must be tackled in South Africa.

Van der Berg and Burger (2003) analyse performance in Western Cape schools using a series of school-level regressions focusing on poverty and school resource variables. They conclude that the resource allocation variables in their data set fail to contribute substantially to the explanation of black and coloured school results (2003: 16). They further conclude that there could be a substantial role for omitted variables such as the availability of teaching resources and managerial (in) efficiency.

Van der Berg (2007) uses South African Grade 6 data from the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) database to model the relationship between performance in mathematics and reading and a range of explanatory variables. This paper appears to have two aims, namely to improve understanding of the role of SES and other factors in determining academic performance and to move us to an understanding of how and under what conditions school resources matter.

Van der Berg employs five statistical models using OLS estimation, quantile regression and hierarchical linear modelling techniques on the full and reduced samples of schools and individuals. In the first two models, clustering is treated as a nuisance and "robust" standard errors are calculated to adjust for clustering, while in the final multilevel model, such clustering is explicitly modelled.

Using the differences in mean scores in mathematics and reading for the fifth quintile and other quintiles, Van der Berg concludes that the data-generating process may be different for historically white and Indian schools (assumed to all lie in quintile 5) and black and coloured schools (assumed to lie in quintiles 1 to 4). This is his rationale for analysing two samples, namely a full sample and a reduced sample that only takes into account data of schools that fall in his quintiles 1 to 4 categories.

Van der Berg finds large differences between the two samples in the size of the estimated coefficients (especially SES) and states that SES ceases to be an important variable in the

Van der Berg correctly concludes that inequality in educational outcomes predates the secondary schooling phase and hence serious remedial action is needed at the primary schooling level.

reduced sample. The conventional resource variables (school buildings, equipment, textbooks) are not statistically significant in either full or reduced samples. Also, the models run on the reduced samples do not appear to have the same explanatory powers as the models for the full sample. Van der Berg attributes the lower explanatory power of the models in the reduced sample to "varying school efficiency". In the multilevel model, most of the results correspond with what was found in the full sample OLS model, with a few exceptions for the mathematics and reading models.

Van der Berg correctly concludes that inequality in educational outcomes predates the secondary schooling phase and hence serious remedial action is needed at the primary schooling level. He also states that there is a relatively large divergence in the ability of schools to convert resources into outcomes.

It is difficult to interpret the findings of the Van der Berg (2007) paper for the following reasons:

- International quantitative research suggests that judgments about school effectiveness cannot be based solely on results, because that presumes a causal relationship between an intervention (school management, good teachers etc.) and results or outcomes. Survey research has not succeeded in isolating this intervention, hence recourse to the next best control, namely entry ability. When prior academic achievement or entry ability is controlled for, many of the large school effects disappear (Yang and Woodhouse, 2001; Raudenbush, 2004), and Van der Berg did not control or have access to this central variable in his study.
- Van der Berg's claim that there is a large divergence in the ability of schools to convert resources into outcomes based on large random effects (principally intercept variance) ignores selection effects evident in parental choices (or lack of choices) for schools and the different intake quality of poor and rich schools. Also, this contradicts his conclusion that resource effects were insignificant for both the rich and poor schools samples.
- The finding that SES does not matter in black and coloured poor schools could be due to the reduced range of the SES variable. Van der Berg does not provide the values of the variables (minimum and maximum) for his reduced sample, thereby not ruling out the possibility that the reduced range of the variable is responsible for the finding of no effect for SES.
- I find his interpretation of the random SES-outcome slope confusing. Mean SES is a fixed effect, and yet it appears in the random parameters of the hierarchical model (tables 9 and 10). Mean SES is incorrectly related to performance when it is, in fact, a fixed school-level variable that is used to explain how the social composition of a school inflates or flattens the SES-outcome slope and/or affects the average performance of schools. This is both a performance and an equity issue, which is not explicitly addressed in his paper.
- Furthermore, the random SES effect (or SES-outcome effect) simply means that this relationship/slope is predicted to randomly vary over schools, yet Van der Berg, in explaining this "random effect", incorrectly states that "many schools deviate from the general pattern of relationships

> **The international literature seems to be characterised by a diversity of contributions across disciplines, and the debates are robust ... The South African literature seems too narrow in scope and methodological diversity, and much more work (and better data) is required before we settle on our own consensual version of what matters, when and where.**

between the school mean SES and individual SES". This is directly related to his interpretation of mean SES as the random effect.

My short review of some of the international literature suggests that debates are no longer focused on whether resources matter; researchers are now interested in estimating models that are more refined and account for methodological problems that reduced the usefulness of earlier work.

Generally, resource effects appear modest (with a few notable exceptions), but none of the reviews imply that these effects are educationally unimportant. Class size has emerged as a powerful explanatory variable, but this requires the research design to be rigorous (longitudinal and controlling for key variables) or endogeneity problems to have been addressed. It is therefore not surprising to see significant effects in these studies. The importance of class size effects is further boosted by the results from the Tennessee randomised trial experiment that found statistically significant effects persisting throughout high schools for students who experienced smaller classes (Raudenbush, 2009).

What is useful about the most recent research is that researchers are thinking about both equity and quality and how resources can be used to firm up this vital connection in education policy, implementation and outcomes. In the language of multilevel models, the search is on for schools with large positive intercepts (quality), but for which the relationship between disadvantage and outcomes is a far flatter relationship/slope (which implies equity).

The international literature seems to be characterised by a diversity of contributions across disciplines, and the debates are robust and, in some instances, clinical in their rejection of poor educational research. In contrast, the South African literature seems too narrow in scope and methodological diversity, and much more work (and better data) is required before we settle on our own consensual version of what matters, when and where.

At a minimum, we need to adopt international best practices in data collection, research designs and analytical rigour. But we need a community of diverse scholars to do this - to enlarge our creative imagination, extend our interventions and encourage novel ways of examining issues, including the resources issue, in South African education.

An exploratory empirical analysis of the school resources and academic outcomes nexus in South African primary schools

Rationale for using multi-level models

The basic ideas behind multilevel statistical regression models concern the way in which clustering and the complex issues of "fixed" and "random" effects are dealt with. Also, the hierarchical nature of data is explicitly taken into account, thus requiring appropriate data at each of the levels of the data hierarchy. If one assumes that individuals who belong to the same cluster (say, schools) are more alike than individuals from another cluster, then the within-cluster correlation should be accounted for. Methodologically, measurement observations (test scores) can also be nested in individuals and similar assumptions about the dependency of separate measurements for the same individual can be made.

Cohen et al. (2003) state that one would expect, in such longitudinal designs, that test measures from any individual would be more correlated with one another than the test scores across individuals. In a cross-sectional design, if there are severe effects of clustering, individuals become copies of each other, thus rendering the idea of independent observations suspect. This then introduces the distinction between nominal and effective sample size. If significance tests and confidence intervals are based on nominal sample size and ignore the effect of clustering, then standard errors (precision of estimates) will be underestimated, leading to optimistic conclusions of an effect where none is actually present in the population from which the sample was drawn.

Snijders and Bosker (1999) argue that one can deal with clustering in two ways, namely as a nuisance (calculating robust standard errors but ignoring the hierarchical nature of the data) or by modelling clustering as an interesting phenomenon in itself.

Kreft and De Leeuw (1998: 11) argue that the distinction between fixed and random effects is useful because it has consequences for the inferences that can be made and for the generalisation of the results. Put simply, fixed effects only allow inferences about the treatments in the experiment (or the estimated effects in the regression model) and are assumed constant and without measurement error. In random effect models, inferences are extended beyond the specific clusters in the sample (say, schools), and Kreft and De Leeuw state that the intention is to generalise to the population of schools and not only to the schools in the treatment (sample study).

In my own understanding, a coefficient is fixed if its effect is assumed to be the same across clusters (the same within- and between-slope is assumed) and it is random when the within-cluster slope is hypothesised to deviate for the weighted average slope that is calculated across the clusters.[6] The deviations (residuals) from the average weighted intercept or slope are further qualified in terms of distributional assumptions, and the effects are considered a random variable with a mean and a standard deviation (variance).

There has been tremendous development in the theory and practice of multilevel models over the past 15 to 20 years. Here is a select list of relevant publications:

- For book-length treatments and peer-reviewed academic articles for the standard multilevel linear and generalised linear model, see Aitkin and Longford (1986); Goldstein (1999); Kreft and De Leeuw (1998); Raudenbush and Bryk (2002); Singer (1998); and Snijders and Bosker, (1999).
- Treatments of the multilevel growth model can be found in Hedeker (2002); Seltzer, Choi and Thum (2003); and Singer and Willet (2003).
- The multilevel factor and latent class models are dealt with in Bartholomew et al. (2008), while excellent (although technical) expositions for the multilevel multiple membership model and crossed random effect models can be found in Fielding and Goldstein (2006); Goldstein, Burgess and McConell (2007); and Raudenbush (1993).

The data and data analytical issues

This paper uses data from the Grade 6 Systemic Evaluation Survey produced by the national Department of Education in 2004. The respondents in the survey were Grade 6 learners, their parents or guardians, educators, principals, district officials and hostel supervisors (HSRC, 2005: 8). In the Systemic Evaluation Survey, province was considered the only stratification variable, while at the school level, the research design was set up so that a random sample of up to 42 learners was considered. Data from 998 schools and 34 015 learners could eventually be used (HSRC, 2005: 14).

In this paper, data for 870 schools and a total of 27 171 learners were used because many of the covariates had missing data. The issue of missing data and the potential bias this may introduce to the results has not been dealt with, but is clearly an important issue for quantitative education researchers. Furthermore, data from the learner, parent and principal modules were used, and data from the district level were used only in calculating the degree of clustering on the outcome variables and the two school resources variables.

The three outcomes have different maximum scores, and to facilitate the interpretation of the coefficients across the three equations, the response variables were transformed using standardised outcomes. Centring affects the interpretation of the intercept and its variance at the school level in a fundamental way, and should therefore be approached with great caution (Enders and Tofighi, 2007).

At the student level, both age and student SES were grand-mean centred. This is appropriate in a context where the interest focuses on the effect of level-2 variables and where level-1 variables are used to adjust the overall effect of the school-level variables. Other individual-level variables were treated as binary variables (gender and attendance at pre-primary school). The reference category for the gender variable was boy learners, and for the pre-primary education it was learners who had not been exposed to any form of pre-school education. I have not heeded advice to use grand-mean centring for the dummy variables, partly because my interests did not focus on offering a substantive interpretation of the intercepts. The standard advice is also to centre level-2 covariates at their grand means, but that has not been done in this paper.

At the school level, I have produced a measure of the variability of intake as per the individual SES of learners. This was aggregated and treated as a school-level variable. An index of school resources was created consisting of variables that measure schools' access to physical and educational resources as well as the total private resources (principally school fees) at the school's disposal. It is important to note that school resources defined this broadly are a much more acceptable definition than simply focusing on isolated school resource variables, because that is not the way schools operate. The second school resource variable used was class size, as reported by the principal. It is possible that this variable was measured with some error, because school officials have been known to inflate estimates of variables that have an impact on their overall resources envelope. Initially, the class size variable was regressed on total enrolment to control for possible endogeneity, and the fitted results were used as a proxy for class size. This is similar to a two-stage least-squares approach (minus the production of appropriate standard errors). However, in the final model, I decided to use the original class size variable as reported by the principals of primary schools.

Other school-level variables included the school SES status (aggregated from individual learners' SES), the urban/rural status of schools, with rural schools as the reference category, the proportion of girl learners in a Grade 6 school and the percentage of learners who indicated participation in the school feeding scheme.

Also measured at the school level were the percentage of learners who indicated that their schools provided for special needs; the ability of a school to accommodate learners, given its maximum capacity; total enrolment; and the total time each school devoted to learning and teaching. The latter variable was treated as a dummy variable, and the reference category was schools that indicated total teaching and learning time of less than 26 hours. In the specification of the model, I was trying to take the advice of Kreft and De Leeuw (1998), who warn that large, complicated multilevel models may be subject to parameter instability. This leaves the model vulnerable to charges of omission bias, but, given that this is an exploratory analysis, further work may potentially increase the subset of relevant covariates.

The absence of prior academic achievement or other proxy measures of learner ability makes it impossible to say anything meaningful about overall school effectiveness, and the reader must bear this in mind when interpreting the results. There is ample evidence internationally that large school effects (essentially residuals from multilevel models) are significantly reduced once this variable is controlled for.

The model

This paper uses a multivariate multilevel model whose purpose is to estimate the effect of independent variables on two or more dependent variables. "Multivariate" here refers to the number of dependent variables to be estimated in a single model. A three-level model will be estimated with measurement on variables (level 1), pupils (level 2) and schools (level 3). Snijders and Bosker (1999) argue that it is not necessary for each individual *I* in each Group *J* to have an observation on each of the *m* variables, but they do caution that missingness at random (MAR) must be assumed (1999: 200). This problem does not affect the present study because the data on the outcomes were nearly complete.

Snijders and Bosker (1999: 201) state that it is possible to analyse all dependent variables separately, but there are a number of reasons why it makes sense to analyse the data jointly as multivariate data:

- Conclusions can be drawn about the correlations between dependent variables, which are made possible by the partitioning of the covariances between dependent variables over the levels of the analysis.
- The tests of specific effects for single dependent variables are more powerful in the multivariate analysis, and testing whether the effect of an independent variable is stronger on dependent variable Y1 and Y2 is only possible by means of a multivariate analysis.
- Finally, if one wishes to carry out a single test of the joint effect of an explanatory variable on multiple dependent variables, then multivariate analysis is required.

The multivariate random intercept model allows the inclusion of fixed effects for the explanatory variables and has an intercept that varies randomly over clustered units (in our case, schools). Following Snijders and Bosker (1999: 201), this model is expressed in the following way:

$$Y_{hij} = \gamma_{0h} + \gamma_{1h} x_{1ij} + \gamma_{ph} x_{pij} + U_{hj} + R_{hij.}$$

(Equation 1)

This model describes the hth dependent variable, the intercept is γ_{0h}, the regression coefficient on X_1 is γ_{1h} etc.,the random part of the intercept in Group J is U_{hj} and the residual at the pupil level is R_{hij}. This is the model that has been employed in the present analysis.

From an estimation and data analysis point of view, estimating more than two dependent variables at the same time requires dummy variables to be worked into the model to indicate which response variable is present. Dummy variables d_1 to d_2 are used to indicate the dependent variables. Following Snijders and Bosker (1999), this can be expressed in the following way:

$$d_{shij} = \begin{cases} 1 & h = s, \\ 0 & h \neq s \end{cases}$$

Following the definition of the dummy variables, each of the explanatory variables in the model (including the intercept and random effects) is multiplied by the dummy variables. This gives the three-level hierarchical linear model, represented in Equation 2:

$$Y_{hij} = \sum_{s=1}^{m} \gamma_{0s} d_{shij} + \sum_{k=1}^{p} \sum_{s=1}^{m} \gamma_{ks} d_{shij} x_{kij} + \sum_{s=1}^{m} U_{sj} d_{shij} + \sum_{s=1}^{m} R_{sij} d_{shij.}$$

(Equation 2)

The coefficients of the dummy variables are assumed to vary randomly across students and schools to obtain the student and school residuals (Steele and Goldstein, 2006). Of course, some researchers would insist that one separate the pupil-level variables from the school-level variables in Equation 2 above so that one has z_{hj} for school-level variables and x_{hij} for pupil-level variables, but the general idea is conveyed well in Equation 2.

Most researchers who model the relationships among a range of explanatory variables and dependent variables start with the multivariate multilevel "empty" model, where no explanatory variables are included. This model allows for the estimation of the means of the response variables and the variance-covariance matrix at the second and third level of the data hierarchy in a three-level model. In this paper, I use the empty model to define the intra-class correlation for each outcome and residual correlations among pairs of outcomes at the two higher levels of the data hierarchy.

$$Y_{hij} = \gamma_{0h} + U_{hj} + R_{hij}$$
$$= \sum_{s=1}^{m} \gamma_{0s} d_{shij} + \sum_{s=1}^{m} U_{sj} d_{shij} + \sum_{s=1}^{m} R_{sij} d_{shij}$$

(Equation 3)

Inevitably, one may want to test for the effect of a random slope. Equation 4 provides the notation for one random slope as there is no need to indicate the precise number of random slopes to be employed in the actual analysis.

$$Y_{hij} = \sum_{s=1}^{m} \gamma_{0s} d_{shij} + \sum_{k=1}^{p} \sum_{s=1}^{m} \gamma_{ks} d_{shij} x_{kij} + \sum_{s=1}^{m} U_{0sj} d_{shij} + \sum_{s=1}^{m} U_{1sj} d_{shij} x_{1ij} + \sum_{s=1}^{m} R_{sij} d_{shij.}$$

(Equation 4)

In the random intercept and slope model, the number of additional variances is determined by the number of random slopes. Naturally, as the number of random slopes increases, the data requirements increase and the variance-covariance matrix of each data level becomes much more complex.

The variance-covariance matrix is represented in the following way. At the school level, it measures the deviation of school means in each of the three outcomes from the overall intercepts predicted for each of the three outcomes. Snijders and Bosker (1999) refer to it as the between-group covariance matrix. The residuals for the three outcomes are assumed to have a multivariate normal distribution with expectation (0) and variance Ω_y.

$$\Omega_v = \begin{pmatrix} \sigma^2_{v0} & & \\ \sigma^2_{v01} & \sigma^2_{v1} & \\ \sigma^2_{v02} & \sigma^2_{v12} & \sigma^2_{v2} \end{pmatrix}$$

(Equation 5)

If a random slope is added to the model, it is represented as a variance on the diagonal, and the covariance with the three intercepts will be on the off-diagonal. In this paper, no random slopes are defined because of my primary focus on understanding the relationship between the two level-2 covariates and the three academic outcomes.

At the student level, the variance-covariance matrix represents the deviation of individual students' scores from their school means. This is why it is more appropriate to refer to this covariance matrix as the *within-group covariance matrix.* In this paper, the student-level covariance matrix assumes the same form as the school-level covariance matrix represented above. However, it is possible to define a much more complicated within-group covariance matrix, especially when one wants to explicitly model a variable-dependent set of residuals. This is not at issue in this paper and treatment of this phenomenon will not be discussed further.

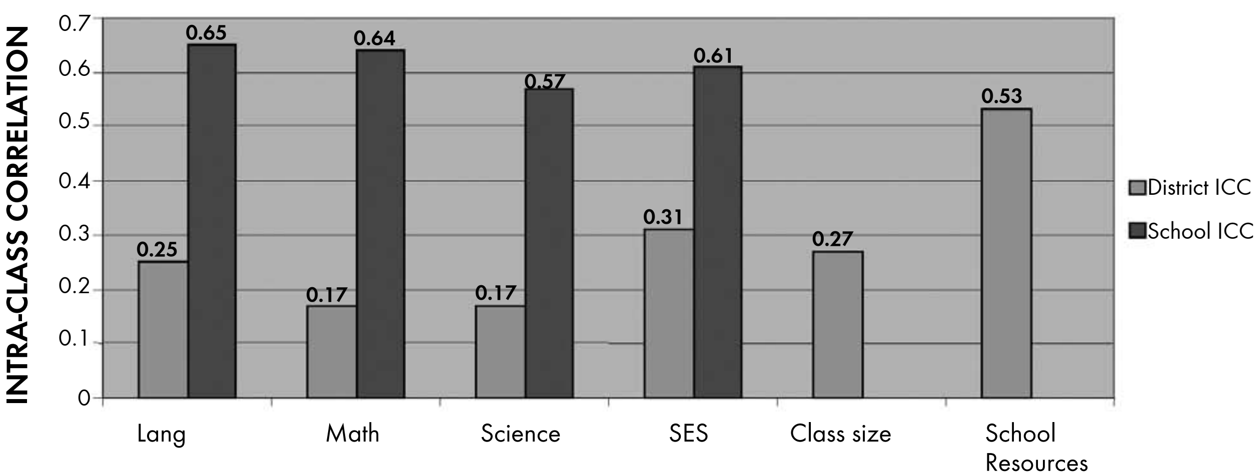
One last word on the variance-covariance matrix at each level: for example, this is particularly useful at level 3 (school-level), where one could calculate the extent to which mean performance in one subject is related to another subject. High residual correlations at this level indicate that mean performance in one subject is strongly related to mean performance in another and also that the unobservable factors that determine the two outcomes may be quite similar. This allows the researcher to further specify the model by adding fixed effects at level 3 (school level) of the multivariate multilevel model. This is also important from a "school effectiveness" point of view, because most school effect studies in South Africa have used overall achievement measures or methods that estimated outcomes in separate models. For the latter models especially, it has not been possible to establish the degree of correlation in performance across subjects, hence the usefulness of the present model.

Discussion of results

Figure 1 provides information on the degree of clustering within schools and districts across the three outcomes and two of the resource variables.

Clustering among students within schools, and among schools within districts, is one of the main reasons why this paper has adopted a multilevel modelling approach. School-level clustering takes into account how the scores of two randomly sampled students

Figure 1: Clustering of outcomes and key covariates at school and district levels



Source: Department of Education, Grade 6 Systemic Evaluation Survey 2004 (own calculations) (language, for example) from a randomly sampled cluster (schools) are correlated. High correlations are indicative of homogeneity within schools and, consequently, a high degree of between-school separation on the measured variable. Van der Berg (2007) has offered a comparative analysis of clustering across country contexts and argued that the degree of clustering within South African primary schools (at Grade 6) is exceptionally high, which is indicative of the large between-school inequalities that exist in South Africa's public schooling system. This is very different from the situation in many developed countries (and some neighbouring SADC countries), where a large percentage of the variability in measured variables exists among students within the same schools. Such schooling systems are therefore more "inclusive" and Willms (2006) presents emerging evidence indicating that such schooling systems tend to perform better on language, mathematics and science outcomes regularly tested for in international assessment tests.

In the Grade 6 data, there is evidence of large within-school homogeneity in the three outcomes, as well as in student SES. In language and mathematics, approximately 65% of the variability in the outcome lies between schools, while 57% of the variability in science outcomes is attributed to the primary schools learners attended. Also, learners who have the same socio-economic backgrounds tend to go to the same schools, thus reinforcing the idea of homogeneous intakes. Hence it is not surprising, given the strong links between SES and performance, that the majority of South African primary and secondary schools are underperforming. Figure 1 also shows that there is a large degree of homogeneity among schools within a district on the school resources variable. Naturally there are districts that are resource-poor as measured by the school resources index variable. Although this district effect has not been taken into account in the present analysis, there is clearly a need to better understand such variation at the district level. However, the class size measure does not appear to have the same high levels of homogeneity among schools within districts, suggesting a more direct link between school policies and provincial policies on class size.

While single measures of clustering are extremely useful, the statistical model employed in this paper allows one to explore the correlations between unobserved determinants (residuals) of the three outcomes at both the student and school levels. Table 1 provides information about the residual correlations among the outcomes at the student and school levels, having only accounted for the group means (the so-called empty model). Subsequently, a number of level-1 and level-2 variables will be added to the model to see if the original residual correlations can be accounted for or "explained".

Residual correlations at the school and student levels measure the degree to which unobserved factors are similar for the respective pair of academic outcomes. At the school level, this says something about the correlation between the average achievement levels for the two outcomes, and at the student level, important information is conveyed about the residual correlations between outcomes among students within a given school. In both instances, modelling the outcomes as a function of explanatory variables should reduce the residual correlations, especially when the model is well specified and variable measurement error is kept to a minimum, and if educationally relevant variables are available and used in the analysis.

The population correlation coefficient for any pair of outcomes can be expressed as follows:

$$\rho(U_{1j,}U_{j2}) = \frac{Cov(U_{1j,}U_{2}j)}{\sqrt{VarU_{1j}} * VarU_{2j}}$$

(Equation 6)

The residual correlation between language and mathematics at the school level is 0.92, while the corresponding correlation between language and science is 0.94. The residual correlation at the school level between mathematics and science is 0.94. The data from the empty model indicates that mean school performance in one outcome is very strongly related to mean performance in one or both of the remaining outcomes. This may mean that the unobservable factors that affect the outcomes could be similar, but this must obviously be explored in the actual statistical model. The results are unsurprising, given the initial discussion of the high levels of homogeneity in the key variables at the student and school levels.

Residual correlations among students within the same school are smaller, but still very high by international standards. The residual correlation between language and mathematics is approximately 0.60, while the correlation between language and science is 0.72. Finally, the residual student correlation between mathematics and science is approximately 0.56. In the case of the student residual correlation between mathematics and science, the empty model suggests that a fair number of mistakes will be made if we try to predict a learner's mathematics score based on knowledge of his science score. Still, it is important to recognise that even at the student level, the residual correlations are high, further reinforcing the point about the homogeneity of students within schools.

Table 2 sets out the results of the final model for the fixed components of the model.

Analysis of the data occurred in five stages. Because of the recurring problems associated with the urban-rural achievement gap, this variable was the first to enter the model. Thereafter, demographic variables (gender, age and SES) were accounted for. Although attendance at pre-primary school is a level-1

TABLE 1: Residual variances and covariances at school and student levels in the empty model

	Language H=1		**Maths h=2**		**Science h=3**		**Covariance (SE)**
Fixed effect	**Parameter**	**SE**	**Parameter**	**SE**	**Parameter**	**SE**	
Intercept	-0.044	0.026	-0.035	0.025	-0.039	0.024	
Random effect							
School level							
Variance	0.649	0.030	0.624	0.029	0.564	0.026	
Cov language and maths							0.586 (0.028)
Cov language and science							0.566 (0.027)
Cov maths and science							0.558 (0.026)
Student level							
Variance	0.334	0.003	0.346	0.003	0.423	0.003	
Cov language and maths							0.203 (0.002)
Cov language and science							0.270 (0.003)
Cov maths and science							0.215 (0.002)

Source: Department of Education, Grade 6 Systemic Evaluation Survey 2004 (own calculations)

TABLE 2: Fixed parameter estimates for the multivariate multilevel model for language, mathematics and science (dependent variables in standard scores)

	Language *h=1*		Maths *h=2*		Science *h=3*	
Fixed effect	**Parameter**	**SE**	**Parameter**	**SE**	**Parameter**	**SE**
Intercept	-1.259	0.111	-0.819	0.122	-0.994	0.123
Gender	0.154	0.007	-0.011	0.007	0.068	0.008
Age	-0.104	0.003	-0.077	0.003	-0.107	0.003
Student SES	0.002	0.001	0.001	0.001	0.001	0.001
Pre-primary education	0.038	0.008	0.046	0.009	0.025	0.009
School SES	0.103	0.007	0.090	0.008	0.092	0.008
Standard deviation of school SES	-0.061	0.020	-0.106	0.022	-0.021	0.022
Urban	0.064	0.079	0.222	0.085	0.069	0.088
% participation in school feeding scheme	-0.002	0.000	-0.003	0.001	-0.002	0.001
% who indicated special needs	-0.001	0.001	-0.001	0.001	-0.002	0.001
26-30 hours teaching and learning time	0.086	0.023	0.076	0.025	0.023	0.025
30 hours plus teaching and learning time	0.061	0.028	0.060	0.030	0.037	0.032
% female students	0.005	0.001	0.004	0.001	0.003	0.001
School resources	0.035	0.005	0.038	0.006	0.027	0.006
Class size	-0.004	0.001	-0.003	0.001	-0.004	0.001
Class size*urban	-0.002	0.002	-0.008	0.002	-0.002	0.002
Class size*school SES	0.008	0.001	0.009	0.002	0.006	0.002

Source: Department of Education, Grade 6 Systemic Evaluation Survey 2004 (own calculations)

variable, it is neither a demographic variable nor a school-context variable, and, for this reason, entered the model after the demographic variables were accounted for. Then followed the school-context variables (standard deviation of student SES, time devoted to teaching and learning, and other relevant school-context variables). A decision was made to exclude the school SES variable because of its large predicted effects. This variable was entered after the other school-context variables to study its impact on the rural-urban performance gap. The school resources index and class size variables were then entered into the model and finally a set of non-linear variables (mostly interactions). Polynomial transformation of the class size variable was attempted, given the so-called threshold finding in class size research, but did not add significantly to the model or change the key parameter estimates of the model. The final model did not specify a random slope because of the primary interest in the effect of school resources on outcomes. To facilitate interpretation of the fixed parameters in the final model, results are reported in the original metric of the response variables, using the standard deviations of the original response variables. (See the appendix for a description of the variables' means and standard deviations.)

Demographic variables

Girl learners are predicted to do better than boy learners in language and science. In language, this advantage translates into approximately 2.9 points, whereas in science, a smaller gain of 0.60 is predicted for girl learners. Girls and boys do not appear to have statistically significant differences in the mathematics outcome. Because it was not a purpose of the present paper to model the determinants of performance per se, no further examination of the gender achievement gap was undertaken, but this is clearly an issue that needs further attention. Older learners in Grade 6 are penalised across the three outcomes, and the disadvantage ranges from two points in mathematics to almost one full point in science. In mathematics, the penalty associated with being an older learner in Grade 6 is half a point. Student SES is widely regarded as one of the most potent predictors of academic performance, yet in the Grade 6 data, only the SES coefficient for language is statistically significant, and it has a negligible effect. For mathematics and science, the SES slope is statistically insignificant. However, the SES measurement used in the analysis does not include parental income and only reflects access to possessions and parental levels of education. Stronger measures may produce the expected results, and hence one should treat the present parameter estimates with some caution.

Importance of pre-primary education

Having attended a pre-primary institution is associated with positive gains across the three outcomes. For language, the gain is approximately 0.70 of a point whereas smaller effects are predicted for mathematics (0.32) and science (0.22). While these are not large effects, it is nonetheless interesting that pre-primary education has this effect, especially given the deplorable state of pre-primary education before and after the first democratic elections. Also, this variable does not measure the quality of pre-primary education, but merely asks whether a learner had attended some form of pre-primary education. Viewed from this perspective and given that this variable does not measure pre-primary quality, but merely quantity, further work needs to go into replicating and understanding the effect found in the sample data.

School context covariates (excluding school SES)

The standard deviation of the SES variable measures the heterogeneity of intake in Grade 6 schools. International research suggests that greater heterogeneity within schools is associated with better performance for the schooling system as a whole. The initial results suggest that an increase in the socio-economic heterogeneity of learners within schools is associated with a penalty in language and mathematics. For language, this penalty is equivalent to about 1.1 points, while for mathematics, it is much smaller, at three-quarters of a raw score point. No statistically significant relationship between the heterogeneity of intake and science performance was found. The school feeding and special needs school variables accounted for negligible negative effects across the three outcomes and do not appear to be powerful predictors of learner performance. Also, the proportion of girl learners at a school has a positive effect, but a very small one, that ranges from 0.001 to 0.003 of a full point. Thus the individual effect of gender does not appear to have any contextual significance.

Not surprisingly, the time devoted to teaching and learning has a positive and significant impact for at least language and mathematics. Schools that devoted between 26 and 30 hours to learning and teaching had an advantage of approximately 1.6 points over schools with fewer hours of learning and teaching time. Also, schools that devoted 30 hours and more to learning and teaching had a net gain of more than 1.1 points over schools that devoted less than 26 hours to learning and teaching. For the mathematics outcome, this advantage is smaller and translates into approximately half of a full point for the two dummy variables.

School socio-economic status and urban/rural status of schools

Before the school SES variable was entered into the model, a positive achievement gain existed for urban schools. However, when the school SES variable was entered, this gain was wiped out for language and science. The slope of urban schools was found to be no different statistically from rural schools for language and science. However, once the full model was specified, urban schools still outperformed rural schools by approximately 1.6 points in mathematics. The present results for language and science suggest that the urban/rural achievement gap has less to do with the inherent characteristics of rural schools than the fact that these schools serve very poor communities. Turning to school SES, this variable was found statistically significant across the three equations, unlike the student version of this variable.

For language, a one-unit increase in school SES is associated with a large gain of 1.9 points, while smaller, yet highly statistically significant, gains are predicted for mathematics and science. The fact that the contextual version of the SES variable appears to play a more important role in determining outcomes is not surprising and is consistent with the strong clustering effects discussed earlier in this paper. Furthermore, such effects could be larger if stronger measures of student SES were available.

School resources variables

Two measures of school resources have been used, namely an index variable summarising the levels of resources at schools' disposal and class size as reported by the principals of the primary schools concerned. The choice of an index variable for school resources was motivated by the fact that school resources come in a package, and it does not make sense to isolate only single variables and study their association with outcomes. The class size variable is subject to inaccuracies, as school officials have a long history of inflating estimates of variables that may impact on the overall school resources envelope. School resources are predicted to have a positive relationship with all three outcomes, although the magnitude of that effect differs across the three outcomes. For language, a one-unit increase in school resources is associated with a gain of almost two-thirds of a point, whereas for mathematics and science, the corresponding gains are approximately a quarter of a point. The class size variable is highly significant across the three outcomes, but its effects are extremely small, and hence one could question the practical significance of this finding. However, this analysis did not consider the endogeneity of the class size variable. This is vital, seeing that class sizes are not randomly allocated within and across schools. Given the initial encouraging finding, more refined models that explicitly account for the potential endogeneity of class size need to be estimated. Based on previous evidence, the positive finding for the school resources and the class size variables are encouraging and must be explored in greater detail in follow-up research.

Class size interactions

Given the centrality of class size in this paper, a range of transformations of the variable was attempted and tested in the model. One of the questions concerned whether the effect of the class size variable was different for urban as opposed to rural schools. No evidence of such differentiation was found for language and science, though a small negative effect was found for mathematics. The slope of the class size and outcome relationship is predicted to be slightly more negative (-0.05 in raw score terms) for urban schools. An interesting finding relates to the relationship between class size and the SES status of the school. The latter variable was recoded to have only two categories, namely a low-SES and a high-SES status. The results indicate that high-SES schools appear to substantially mute the negative effect of class size and that the sign of this relationship may even change in high-SES schools. This may be due to the extra capacity high-SES schools have

TABLE 3: Random parameter estimates for the multivariate multilevel model for language, mathematics and science

	Language *h=1*		**Maths** *h=2*		**Science** *h=3*		**Covariance** *(SE)*
Random effect	Parameter	SE	Parameter	SE	Parameter	SE	
School level							
Variance	0.131	0.007	0.162	0.008	0.162	0.009	
Cov language and maths							0.101 (0.007)
Cov language and science							0.113 (0.007)
Cov maths and science							0.131 (0.008)
Student level							
Variance	0.303	0.003	0.338	0.003	0.396	0.003	
Cov language and maths							0.188 (0.002)
Cov language and science							0.244 (0.003)
Cov maths and science							0.198 (0.003)

Source: Department of Education, Grade 6 Systemic Evaluation Survey 2004 (own calculations)

to accommodate larger class sizes, which is an option that is not available to low-SES schools, or something over which they do not have much control.

The results for the fixed parameters of the model having been briefly discussed, table 3 provides information about the random parameters after the final model has been fitted.

Having fitted the final model, we can now compare the residuals at the school and student levels with those of the empty model. At the school level, the residual correlation between language and mathematics has been reduced from 0.92 to 0.69. In the context of the limited model, this reduction is substantial and reflects on the strong role of level-2 variables in the final model.

The residual correlation between language and science has been reduced from 0.94 in the empty model to 0.77 in the fitted model, while the residual association between mathematics and science has been reduced by 14% from 0.94 to 0.81.

In spite of the fitted model, the residual correlations are still high, suggesting that mean performance in one outcome can be predicted with a reasonable amount of confidence from results in the other outcomes. Of course, the absence of a measure of academic ability means that residual correlations could be lowered even more, resulting in a much more powerful explanatory model.

At the student level, because of the small number of level-1 variables that were fitted to the model, residual correlations appear very similar in the fitted and the empty models. This is not surprising, as the main research question concerned the relationship between two level-2 variables and the three outcomes and adjusting for a small number of level-1 covariates.

Concluding remarks

South African primary schools are seriously segregated, and this has enormous implications for how the performance deficits are addressed in historically disadvantaged schools.

Segregation is most evident in the similarity of the socio-economic backgrounds of learners going to the same schools. In the data used in this paper, the intra-class correlation is as high as 0.61, which means the post-apartheid period has entrenched and possibly enlarged socio-economic segregation in primary schools. If one assumes that such patterns are undisturbed for secondary school learners, then one should be extra cautious in drawing causal conclusions about the effectiveness of schools or their differential ability to use scarce and limited state resources. Racial mobility in the post-apartheid period appears to have benefited only a small section of the historically disadvantaged communities, leaving the socio-demographics of schooling largely unchanged.

In this segregated context, heterogeneity within schools, defined as accommodating learners from different socio-economic backgrounds in the same school, appears to have a negative effect on overall performance in language and mathematics. This may reflect in part the strong drive of most middle-class schools to exclude learners from poor backgrounds because of the unresolved issues around state subsidisation of such learners or the fact that the South African schooling system does not tolerate a broader notion of inclusiveness. Furthermore, rural schools, which have a large number of low socio-economic schools, continue to underperform relative to their urban counterparts. However, once the average SES of schools is controlled for, the apparent performance differences between rural and urban schools disappear. This suggests that part of the reason for the recurring underperformance of rural schools is the fact that a very large percentage of such schools are low-SES schools. The evidence in this paper draws one closer to an economic interpretation of the present state of rural schools instead of attributing the poor results of these schools to some inherent characteristics of rural schools and rural areas. Girl learners are predicted to do much better than boys in language and science, but no evidence has been found for gender differences in mathematics performance. This is a relatively unexplored area in South African quantitative education research and more attention needs to be devoted to the emerging boy-learner problem. Older learners in Grade 6 are predicted to do much worse than their age-appropriate fellow learners, and this would therefore be a serious problem for provinces that have a large percentage of older learners in primary schools.

Attendance at some pre-schooling institution is associated with positive benefits across the three outcomes. This is remarkable, given the poor state of early childhood education in the apartheid and post-apartheid periods and the fact that this variable only measured learners' access to pre-primary education. The time devoted to learning and teaching is a significant predictor of learner and school performance. This needs to be explored in greater depth, especially with data that would allow one to understand what actually happens in classrooms. The data used in this paper does not allow one to match teachers to learners, and thus a vital component of the education process remains hidden from analysis. It is entirely feasible that there are other intervening variables that would allow researchers to better explain this connection and to identify factors that would maximise the gains from longer learning and teaching times.

School resources matter, even though its estimated effects are still relatively modest. This is applicable to the index of school resources and the class size variable used in the analysis. With the former, resources gains range from a quarter of point to almost two-thirds of a point. The effects for class sizes across the three outcomes are much smaller, but the fact that these effects are highly statistically significant suggests we need more sophisticated analyses to unpack the resource effects. Recent advances in multilevel modelling now make this possible, and further research should specifically examine potential endogeneity problems in the estimation of school resources. The effect of class size does not appear to be different for urban as opposed to rural schools, but high-SES schools appear to nullify the effect of large classes. This is clearly another area of research that must be pursued, because it is important to know why larger classes appear to affect high-SES schools less than poor primary schools.

In studies in the UK, researchers were able to make a powerful argument for the effect of resources on outcomes, especially for schools serving poor learner communities. Such funding patterns appear relatively stable, and hence the statistical models researchers used were entirely appropriate for that context. In short, they were able to argue for larger portions of funding to be given to learners from poor communities. Because the resource allocation process and the academic outcome process were related, researchers were able to adjust their corresponding estimates of the relationship between resources and outcomes. This was done to avoid a spurious negative correlation between resources and outcomes.

In South Africa, although an increasingly larger share of state resources is given to poor schools, this process was and is at best uneven and powerfully dependent on the availability of resources (Wildeman, 2008). It is clear from the resources index that a very small percentage of South African primary schools can be classified as high-SES schools, and there is a large performance gap between low-SES and high-SES schools. Thus it is premature to draw definitive conclusions about the relationship between resources and outcomes for poor schools when the resource transformation process is far from complete and downright unequal.

Resources are important, but their availability and efficient use will not in themselves solve the education quality gap in South Africa. Two important factors counter an exclusively resource approach, namely the high between-school socio-economic segregation and the fact that we know so little about what happens in primary school classrooms. Better resource conditions will undoubtedly help, but we now need to shine the investigative torch into the classroom, which is, after all, the immediate point of contact for learning and development.

The fact that the present data survey does not allow one to match teachers to classes and learners is symptomatic of an overall avoidance syndrome regarding classroom teaching. This problem involves all stakeholders and must be addressed at all levels of the system. Without such actions, resource benefits will continue to accrue to very specific groups of learners and schools and the desired outcomes will remain glaringly unequal.

References

Aitkin, M. and Longford, N. 1986. "Statistical Modelling Issues in School Effectiveness Studies", *Journal of the Royal Statistical Society*, Series A, 149(1): 1-43.

Bartholomew, D.J., Steele, F., Moustaki, I. and Galbraith, J.I. 2008. *Analysis of Multivariate Social Science Data*, 2nd Edition, London: CRC Press.

Blatchford, P., Goldstein, H., Martin, C. and Browne, W. 2002. "A Study of Class Size Effects in English School Reception Year Classes", *British Educational Research Journal*, 28: 169-85.

Bryk, A.S. and Schneider, B. 2002. *Trust in Schools: A Core Resource for School Improvement*, New York: Russell Sage Foundation.

Case, A. and Deaton, A. 1999. "School Inputs and Educational Outcomes in South Africa", *Quarterly Journal of Economics*, 114(3): 1047-84.

Cohen, J., Cohen, P., West, S. J. and Aiken, L. S. 2003. *Applied Multiple Regression/Correlation for the Behavioral Sciences,* 3rd edition, Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Crouch, L. 2005. "South Africa Equity and Quality Reforms", *Journal of Education for International Development*, 1(1). Retrieved December 2008 from http://www.equip123.net/JEID/articles/1/1-2.pdf.

Crouch, L. and Mabogoane, T. 1998a. "No Magic Bullets, Just Tracer Bullets: The Role of Learning Resources, Social Advantage and Education Management in Improving the Performance of South African Schools", unpublished paper.

Crouch, L. and Mabogoane, T. 1998b. "When the Residuals Matter More than the Coefficients: An Educational Perspective", *Journal of Studies in Economics and Econometrics*, 22(2): 1-13.

Enders, C. and Tofighi, D. 2007. "Centering Predictor Variables in Cross-Sectional Multilevel Models: A New Look at an Old Issue", *Psychological Methods*, 12(2): 121-38.

Fielding, A. and Goldstein, H. 2006. "Cross-Classified and Multiple Membership Structures in Multilevel Models: An Introduction and Review", *Research Report* No 791, London: Department for Education and Skills.

Goldhaber, D. D. and Brewer, D. J. 1997. "Why Don't Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables on Educational Productivity", *Journal of Human Resources*, 32(2): 505-23.

Goldstein, H. 1981. "Models for Reality: New Approaches to the Understanding of Educational Processes", professorial lecture given at the Institute of Education, University of London, 1 July 1998. Retrieved from http://www.mlwin.com/hgpersonal/models%20for%20reality.pdf.

Goldstein, H. 1999. *Multilevel Statistical Models*, 2nd edition, London: Arnold.

Goldstein, H., Burgess, S., and McConell, B. 2007. "Modelling the Effect of Pupil Mobility on School Differences in Educational Achievement", *Journal of the Royal Statistical Society*, Series A, 170 (4): 941-54.

Greenwald, R., Hedges, L. V. and Laine, R. D. 1996. "The Effect of School Resources on Student Achievement", *Review of Educational Research,* 66(3): 361-96.

Hanushek, E. A. 1986. "The Economics of Schooling: Production and Efficiency in Public Schools", *Journal of Economic Literature*, XXIV (September): 1141-77.

Hedeker, D. 2002. "An Introduction to Growth Modeling", in D. Kaplan (ed.)

Handbook of Quantitative Methodology for the Social Sciences, Thousand Oaks, CA: Sage Publications.

Hedges, L. V., Laine, R. D. and Greenwald, R. 1994a. "Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes", *Educational Researcher*, 23(3): 5-14.

Hedges, L. V., Laine, R. D. and Greenwald, R. 1994b. "Money Does Matter Somewhere: A Reply to Hanushek", *Educational Researcher*, 23(4): 9-10.

HSRC, 2005. *Grade 6 Systemic Evaluation: National*, copy received in personal correspondence with Dr Anil Kanjee, Research Director at the Human Sciences Research Council.

Hunt, M. 1997. *How Science Takes Stock: The Story of Meta-Analysis*, New York: Russell Sage Foundation.

Kreft, I. and De Leeuw, J. 1998. *Introducing Multilevel Modelling*, London: Sage Publications.

Levacic, R., Jenkins, A., Vignoles, A., Steele, F. and Allen, R. 2005. "Estimating the Relationship Between School Resources and Pupil Attainment at Key Stage 3", *Research Brief No RB679*, London: Department for Education and Skills.

Levacic, R. and Vignoles, A. 2002. "Researching the Links Between School Resources and Student Outcomes in the UK: A Review of Issues and Evidence", *Education Economics*, 10(3): 313-31.

Murnane, R. J. 1991. "Interpreting the Evidence on 'Does Money Matter?'", *Harvard Journal of Legislation*, 28(2): 457-64.

Rasbash, J., Steele, F., Browne, W. and Prosser, B. 2005. *A User's Guide to MLwiN Version 2.0*, Bristol: Centre for Multilevel Modelling, University of Bristol.

Raudenbush, S. W. 1993. "A Crossed Random Effects Model for Unbalanced Data with Applications in Cross Sectional and Longitudinal Research", *Journal of Educational Statistics*, 18: 321-49.

Raudenbush, S. W. 2004. "Schooling, Statistics, And Poverty: Can We Measure School Improvement?", Ninth Annual William H. Angoff Memorial Lecture, presented at Educational Testing Service, Princeton, NJ, 1 April 2004.

Raudenbush, S. W. 2009. "The Brown Legacy and the O'Connor Challenge: Transforming Schools in the Images of Children's Potential", Fifth Annual AERA Browne Lecture in Education Research (paper received in personal correspondence with the author).

Raudenbush, S. and Bryk, A. S. 1986. "A Hierarchical Model for Studying School Effects", *Sociology of Education*, 59 (January): 1-17.

Raudenbush, S. W. and Bryk, A. S. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd edition, Newbury Park, CA: Sage Publications.

Robinson, W. S. 1950. "Ecological Correlations and the Behavior of Individuals", *American Sociological Review*, 15(2): 351-57.

Seltzer, M., Choi, K. and Thum, Y. M. 2003. "Examining Relationships Between Where Students Start and How Rapidly they Progress: Using New Developments in Growth Modeling to Gain Insight into the Distribution of Achievement Within Schools", *Educational Evaluation and Policy Analysis*, 25(3): 263-86.

Singer, J. D. 1998. "Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models", *Journal of Educational and Behavioral Research*, 24(4): 323-55.

Singer, J. D. and Willet, J. B. 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*, New York: Oxford University Press.

Snijders, T. and Bosker, R. 1999. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling*, London: Sage Publications.

Steele, F. and Goldstein, H. 2006. "Multilevel Models in Psychometrics", in C. R. Rao and S. Sinharay (eds.) *Handbook of Statistics*, Amsterdam: Elsevier.

Steele, F., Vignoles, A. and Jenkins, A. 2007. "The Impact of School Resources on Pupil Attainment: A Multilevel Simultaneous Equation Modelling Approach", *Journal of the Royal Statistical Society Series A* (Statistics in Society), 170(3): 801-24.

Van der Berg, S. 2007. "How Effective Are Poor Schools? Poverty and Educational Outcomes in South Africa", paper presented at the CEGE Research Colloquium, University of Göttingen, Germany, November.

Van der Berg, S. and Burger, R. 2003. "Education and Socio-Economic Differentials: A Study of School Performance in the Western Cape", Working Paper 03/73, Cape Town: Development Policy Research Unit, University of Cape Town.

Wildeman, R. A. 2008. "Reviewing Eight Years of the Implementation of the School Funding Norms, 2000 to 2008", research paper for the Economic Governance Programme of Idasa, Pretoria: Idasa.

Willms, J. D. 2006. "Learning Divides: Ten Policy Questions about the Performance and Equity of Schools and Schooling Systems", report produced for UNESCO Institute for Statistics, Montreal.

Woodhouse, G. and Goldstein, H. 1988. "Educational Performance Indicators and LEA League Tables", *Oxford Review of Education*, 14(3): 301-20.

Yang, M. and Woodhouse, G. 2001. "Progress from GCSE to A and AS level: Institutional and Gender Differences, and Trends Over Time", *British Educational Research Journal*, 27(3): 245-67.

Yang, M., Goldstein, H., Browne, W. and Woodhouse, G. 2002. "Multivariate Multilevel Analysis of Examination Results", *Journal of the Royal Statistical Society*, Series A, 165(1): 137-53.

Appendix

Description of the main variables used in the analysis

	n	Mean	Std Dev	Min	Max
Pupil-level variables					
Language	27171	29.31	18.55	0	76
Mathematics	27171	11.05	7.13	0	40
Science	27171	19.74	8.90	0	46
Gender	27171	0.49	0.50	0	1
Age	27171	0	1.37	-2.53	4.47
Student SES	27171	0	5.10	-9.49	12.51
Pre-primary education	27171	.72	0.45	0	1
School-level variables					
School SES	807	10.52	4.09	2.36	23
School std dev SES	807	3.03	0.69	0	6.23
Time for teaching and learning	807	0.93	0.66	0	2
% girls in school	807	49.53	10.16	0	100
% learners who participate in school feeding	807	65.67	32.65	0	100
% learners special needs	807	74.15	24.79	0	100
Index of school resources	807	8.64	4.35	1	20
Class size	807	40.87	12.90	5	117
Urban	807	0.40	0.49	0	1
School quintile	807	0.20	0.40	0	1

Source: Department of Education, Grade 6 Systemic Evaluation Survey 2004 (own transformations of the explanatory variables)

Endnotes

1 In the monitoring and evaluation literature, outputs and outcomes are very different entities. Although the education literature distinguishes between outputs (immediate results of the system such as performance in tests) and outcomes (long-range results including employment and earnings), this paper defines outcomes as the immediate results of the system, which are mainly results of standardised cognitive achievement tests.

2 For an illuminating discussion on the interchange between Hanushek and the Hedges group, see the book by Morton Hunt (1997), *How Science Takes Stock: The Story of Meta-Analysis*. Apart from documenting this interesting and informative discussion, Hunt also discusses the use of meta-analysis in fields such as psychotherapy and medicine and provides an historical account of the beginnings of meta-analysis as a synthesis research method.

3 Snijders and Bosker (1999: 36) note that Fisher's combination test makes it possible to include p-values of studies that do not use the same operationalisation of variables or research designs. They state that this hypothesis can be tested by calculating minus twice the sum of the natural logarithms of the p-values, represented thus:

$$\chi^2 = -2\sum_{j=1}^{N} \ln(p_j)$$

The authors note because of the shape of the log distribution, this combined statistic will have a large value if at least one of the p-values is very small.

4 The finding that children from poor families who have above-average academic abilities may benefit from better school resources highlights just how important it is to control for prior academic achievement or other similar levels of ability. Also, in the South African context, such learners often go to schools where the average intake quality must be poor, and it must be a stated aim of quantitative research to better understand how such learners are affected by the social composition of the schools they attend. However, none of the available South African surveys allows one to explicitly test and model such relationships.

5 Kreft and De Leeuw (1998) make the important point that the aggregated version of a variable and its manifestation as an individual-level variable ("education" for example) may not necessarily refer to or measure the same concept. They provide an example of how education in an aggregated model has a negative relationship to income, while the individual study shows a positive effect of education on income. Also see Robinson (1950).

6 Kreft and De Leeuw (1998) also refer to the distinction between fixed and random variables. Cohen et al. (2003) state that random variables are those variables whose values are selected at random from a probability distribution. The error terms and the dependent variable(s) in regression models are random variables. In contrast, the independent variables are assumed fixed, although this assumption is regularly violated in applications of OLS regression.

This study probes the connection between resources and outcomes in South African public schools. It aims to develop a deeper understanding of the effects of socio-economic segregation and inadequate resources on the lives of poor learners.

Education is a potent tool in the fight against poverty and other forms of inequality, but to be effective, changes in the schooling system and the way society responds to the challenge of poverty are needed. This study argues that the problems of poor academic performance in public schools cannot simply be attributed to poor teaching, management and the inefficient use of resources. We need to face up to the broader challenges of education for the poor.

This publication is the first in a series of research projects by the Economic Governance Programme of democracy institute Idasa, which locates schools within society and reflects society through the schools and schooling systems it produces.

idasa
AN AFRICAN DEMOCRACY INSTITUTE