Explaining differences in mathematics and reading achievement among Standard 6 pupils in Kenya: Emerging policy issues

Thuku, F.W.
(Egerton University, Kenya)

Hungi, N.
(Australian Council for Educational Research, Sydney)

Abstract

This study employed a multilevel analyses procedure to examine pupil, class and school levels factors that influenced achievement in mathematics and reading of Standard 6 primary school pupils in Kenya as well as to examine performance of schools across provinces in Kenya. The data for this study were collected as part of a major project (known as Southern Africa Consortium for Monitoring Educational Quality, SACMEQ) that sought to examine the quality of education offered in primary schools in Kenya as well as in primary schools in another 15 Southern African countries.

Results show that pupil's age, pupil's home background and pupil-teacher ratio are important factors in the prediction of achievement in mathematics and reading at Standard 6 level in Kenya. In addition, the results show that, when value-added approaches are used to examine school effectiveness, there are no significant differences in performance of primary schools across provinces in Kenya in mathematics and reading. The exceptions here were schools in Nairobi Province which appeared to perform significantly above average in reading but not in mathematics. Policy implications of the findings are outlined.

1 Introduction

There were three purposes of the current study. The first purpose, which was the main purpose of this study, was to identify the key pupil-, class- and school-related factors that influenced achievement in mathematics and reading among Standard 6 pupils in Kenya. The second purpose was to develop multilevel models that could be used to explain some of the variance in mathematics and reading achievement among Standard 6 pupils in Kenya. The third purpose was to examine school effects across provinces in Kenya based on the multilevel models developed in this study.

In order to achieve the above purposes, a three-level model was hypothesized and the data analyzed using multilevel analysis procedures for each of the two outcome measures (mathematics and reading). These data were collected from 3,299 pupils in 320 classes in 185 schools in eight provinces in Kenya as part of the Southern Africa Consortium for Monitoring Educational Quality (SACMEQ) II project in 2002. The SACMEQ II project sought to examine the quality of the education provided in primary schools in Kenya and another 15 Southern African countries.

The structure of this paper is as follows. A section is included in which the techniques used to analyze the data in this study are introduced followed by a section in which the data involved are described. After that, two sections are provided in which the hypothesized multilevel models and the multilevel analyses are described. Finally, sections are included in which results of the analyses are presented and interpretations of the results and their implications for policy are discussed.

2 Approaches employed

The first part of this section introduces the multilevel analysis procedure used to tease out factors influencing pupil achievement in this study while the second part introduces the technique used to estimate school effects in this study.

2.1 Multilevel approach

It is generally accepted that when dealing with multilevel data such as the data in this study, the appropriate procedure is to formulate multilevel models, "which enable the testing of hypotheses about effects occurring within each level and the interrelations among them" (Raudenbush and Bryk, 1994, p. 2590). Testing of hypotheses in multilevel models can be carried out using multilevel data analyses software such as HLM5 for Windows (Raudenbush, Bryk, Cheong, and Congdon, 2000*a*). The HLM program was initially developed to find a solution for the methodological weakness of educational research studies during the early 1980s, which was the failure of many analytical studies to attend to the hierarchical, multilevel character of much of educational field research data (Bryk and Raudenbush, 1992). This failure came from the fact that "the traditional linear models used by most researchers require the assumption that subjects respond independently to educational programs" (Raudenbush and Bryk, 1994, p. 2590). In practice, most educational research studies select students as a sample who are nested within classrooms, and the classrooms are in turn nested within schools, schools within geographical locations, states, or countries. In this situation, the students selected in the study are not independent, but rather nested within organizational units and ignoring this fact results in the problems of "aggregation bias and misestimated precision" (Raudenbush and Bryk, 1994, p. 2590).

The multilevel technique employed in this study has been used in other studies to examine factors influencing pupil achievement in several developing countries. For example, Willms and Somers (2001) used hierarchical models to examine socioeconomic factors and other factors influencing mathematics and language achievement among grades 3 and 4 pupils from 13 Latin American countries. In addition, a study by Howie (2002) successively employed this multilevel technique to tease out language and other factors influencing student achievement in mathematics in South Africa using data that were collected as part of the TIMSS-R study.

2.2 Value-added approach

Comparison of schools using unadjusted average test scores of pupils is argued to be inappropriate because pupils are not allocated to schools at random, neither are schools located in areas with similar neighbourhood characteristics, nor are all schools of the same size. As a result, schools differ in their pupil intakes as assessed by such characteristics as prior achievement, home background and gender. Many past studies have ascertained that pupil-level factors such as level of achievement at entry, family background, and school-level factors such as locality and school type may affect pupils' level of achievement (e.g. Husén, 1967; Comber and Keeves, 1973; Keeves, 1975; Postlethwaite and Wiley, 1991).

Consequently, researchers argue that it would be misleading to compare schools using average scores without adjusting the scores for the differences between schools in their pupil intake (e.g. McPherson, 1993; Yang, Goldstein, Rath and Hill, 1999). The adjustment of the scores for the differences between schools enables the boost (called 'value added') that each school provides to its pupils' achievement to be more apparent. In other words, value added measures are intended to allow fairer comparison between schools.

The current study employed the theoretical and statistical concepts of the model used by Willms and Raudenbush (1989; pp.212-14) and Raudenbush and Willms (1995; pp.313-19) to estimate school effects. This model hypothesized that a student's academic outcome (Y) was influenced by three general factors; the student background characteristics (S), school context (C), and school policies and practices (P). The linear relationship between the components involved in this model can be written in equation format as follows.

$$Y_{ik} = \gamma_{00} + C_k + P_k + u_{0k} + S_{ik} + r_{ik}$$ **Equation 1**

where:

Y_{ik} is the outcome score for student *i* in school *k*;

γ_{00} is the grand mean;

S_{ik} is the contribution of the background characteristics of student *i* in school *k* (e.g. gender, age and home background);

r_{ik} is a random error or ‘student effect’, that is, the deviation of the student mean from the school mean;

C_k is the contribution of school context (e.g. aspects of school composition such as the average socioeconomic level of the students in the school);

P_k is the effect of school policy and practice (for example, aspects of school leadership, use of resources, curricular content, and classroom instructional strategies);

u_{0k} is a school-level residual also called a random ‘school effect’, that is, the deviation of the school mean from the grand mean (γ_{00}) and it “represents the unique contribution of each school that is not explained by school-level variables in the model” (Willms and Raudenbush, 1989; p. 212).

The indices *i* and *k* denote students and schools where there are *i* = 1, 2, . . . , n_j students within school k; and *k* = 1, 2, . . . , K schools.

The authors illustrated how the above model could be used to estimate two types of school effects.

The first is Type A, defined as:

$$A_k = C_k + P_k + u_{0k} \qquad \textbf{Equation 2}$$

where:

A_k is the Type A effect of school *k*; and

C_k, P_k and u_{0k} have the same meaning as described above.

Thus, Type A effect included the effects of school context, policy and practice, and therefore, the authors said it was an indicator of how well a student of average background characteristics would perform in school *k*, relative to the grand mean. Raudenbush and Willms (1995) argued that the Type A effect was the effect parents generally considered when choosing one of the *K* schools for their child. In addition, they argued that it would clearly be unfair to reward those who worked in the school on the basis of Type A because the school staff was only partly responsible for those effects.

The second is the Type B effect, defined as:

$$B_k = P_k + u_{0k} \qquad \textbf{Equation 3}$$

where:

B_k is Type B effect of school *k*; and

P_k and u_{0k} have the same meaning as described above.

Thus, Type B effect included only the effects of school policy and practice, and therefore, it was an indicator of how well a particular school performed relative to other schools with similar student intake and context. Importantly, Type B excluded factors that lay outside the control of those who worked in the school. Consequently, Raudenbush and Willms (1995; p.310) argued that Type B effect “is the effect school officials consider when evaluating the performance of those who work in the schools”.

3 The Data

It has been mentioned in the introduction that the data for this study were collected as part of SACMEQ II project. The outcome variables of interest in the SACMEQ II project were pupil scores (on Rasch scales) in mathematics and reading tests at Standard 6.

Apart from achievement scores in mathematics and reading, a wide range of information about these pupils, their teachers, their school heads and characteristics of their schools were collected. The variables examined in this study were those variables identified as potential predictors of academic achievement following preliminary analyses, sound reasoning and research findings from studies in other countries. The variables examined in the current study are outlined in a separate section below together with the hypothesized models.

An interesting aspect of the SACMEQ II study was that the teachers were also assessed in mathematics and reading. The teacher and pupil tests used different sets of items but some common items were included in both tests for purposes of equating. The teacher tests were designed to be more difficult than the pupil tests. Teachers' scores in these tests were used as predictors of pupil achievement in mathematics and reading models hypothesized in this study.

4 Hypothesized models

In the multilevel analyses reported in this paper, two separate three-level models were hypothesized and examined, one for factors influencing achievement in mathematics and the other for factors influencing achievement in reading. The outcome variables of interest in these models were pupils' scores in mathematics (ZMALOCP) and reading (ZRALOCP) tests, both with a mean of 500 and a standard deviation of 100. The hierarchical structures of these models were pupils at level-1, class at level-2 and school at level-3. In other words, pupils were nested within classes and classes nested within schools.

In this three-level model, the number of variables that were initially hypothesized to directly influence achievement at the pupil, class and school levels were 15, 25 and 26 variables respectively. In general, there were two types of variables examined for inclusion in the model at the class and school levels. The first types of variables were pupil-related variables (i.e. school context) and these variables were constructed by aggregating the pupil-level data. For example, pupil-level data on the variable ‘Days absent’ were aggregated at the class level in order to construct the variable ‘Average days absent’ at the class level and data on this variable were aggregated at the school level in order to construct the variable ‘Average days absent’ at the school level. The second types of variables were pupil-free variables and these variables were constructed from school characteristic data (e.g. School location), teachers’ characteristics data (e.g. Average teacher training) and community characteristics data (e.g. Parental contribution towards school development).

The names and codes of all the predictor variables tested (whether significant or not) for inclusion at each level of the three-level hierarchical model are provided in *Table 1*. All variables for which data were available for testing are listed in *Table 1*, to show the very extensive range of possible effects that were examined, rather than to provide information only on those that were statistically significant. The lack of statistical significance can sometimes be of great interest in the development or modification of policy.

<Insert Table 1 about here>

5 Analyses

There are two parts included in this section. The analyses undertaken to tease out the factors influencing mathematics and reading achievement among Standard 6 pupils in Kenya are described in the first part and the analyses carried out to examine school effects across provinces in Kenya are described in the second part.

5.1 Teasing out achievement factors

A preliminary task in HLM analyses was to build two sufficient statistics matrix (SSM) files, one for mathematics and the other for reading. No pupils, classes or schools were dropped because of insufficient data in the construction of these SSM files. Therefore, the Ns in these SSM files remained as they were in the original data files; that is, 3,299 for pupils, 320 for classes and 185 for schools. Weighting (with sampling weights calculated to cater for the design of this study) was undertaken in the construction of these SSM files.

The first step in HLM analyses was to run the so-called 'null models' in order to obtain the amounts of variance available to be explained at each level of the hierarchy (Bryk and Raudenbush, 1992). The null models contained only the dependent variables (ZMALOCP for mathematics and ZRALOCP for reading) and no predictor variables were specified at the pupil, class and school levels.

The null model can be stated in equation form as follows.

Level-1 model

$$Y_{ijk} = \pi_{0jk} + e_{ijk}$$

Level-2 model

$$\pi_{0jk} = \beta_{00k} + r_{0jk}$$

Level-3 model

$$\beta_{00k} = \gamma_{000} + u_{00k}$$

Equation 4

where:

Y_{ijk} is the mathematics (or reading) achievement of pupil *i* in class *j* in school *k*.

π_{0jk} is the mean achievement of class *j* in school *k*.

e_{ijk} is 'pupil effect', that is, the deviation of the pupil mean from the school mean.

β_{00k} is the mean achievement of school *k*.

r_{0jk} is 'class effect', that is, the deviation of a class mean from the school mean.

γ_{000} is the grand mean.

u_{00k} is 'school effect', that is, the deviation of school mean from the grand mean.

The indices *i*, *j* and *k* denote pupils, classes and schools. There are

$i = 1, 2, \ldots, n_{jk}$ pupils within class *j* in school *k*;

$j = 1, 2, \ldots, J_k$ classes within school *k*; and

$k = 1, 2, \ldots, K$ schools (in this study, K = 185).

At this stage, the 'basic specification' option in HLM5 was used to generate school-level residual files. The school-level empirical Bayes (EB) residuals obtained from the null model (to be called 'raw effects') are used later in this study in the examination of school effects across provinces in Kenya.

The second step undertaken was to build up the pupil-level model or the so-called 'unconditional' model at level-1. This involved adding pupil-level predictors to the above model (Equation 4), but without entering predictors at any of the other levels of the hierarchy. At this stage, a 'step-up' approach (Bryk and Raudenbush, 1992) was followed to examine which of the 15 pupil-level variables (listed in *Table 1*) had a significant (at $p \leq 0.05$) influence on the outcome variables, ZMALOCP and ZRALOCP. Bryk and Raudenbush (1992) recommended the step-up approach for inclusion of variables into the model to the alternative approach often referred to as 'working-backward' where all the possible predictors were included in the model and then the non-significant variables were progressively eliminated from the model.

The third step in the HLM analyses involved adding level-2 predictors into the model using the step-up strategy mentioned above. The level-2 exploratory analysis sub-routine available in HLM5 was employed for examining the potentially significant level-2 predictors (as shown in the output) in successive HLM runs.

The final step involved building up the model to the school-level through adding the significant school-level predictor variables into the model using the level-3 exploratory analysis sub-routine and the step-up strategy.

In the analyses described above, coefficients of a variable were specified as 'fixed' at a particular level of the hierarchy if the reliability estimate of the variable dropped below the value of 0.05 at that level (see Raudenbush and Bryk, 2002, p.125). Specifying a variable as fixed constrained its slope or intercept to be the same across all the level-2 and level-3 units. The alternative was to specify a variable as 'random', which allowed the slopes and intercepts to vary among the level-2 and level-3 units (Raudenbush, Bryk, Cheong, and Congdon, 2000*b*).

The final models for mathematics and reading are presented below (Equations 5 and 6 respectively).

Mathematics

Level-1 model

$$
\begin{aligned}
[ZMALOCP]_{ijk} = {} & \pi_{0jk} + \pi_{1jk}(AGE)_{ijk} + \pi_{2jk}(SEX)_{ijk} + \pi_{3jk}(ENGLISH)_{ijk} + \\
& \pi_{4jk}(REPEAT)_{ijk} + \pi_{5jk}(MHWORKC)_{ijk} + \pi_{6jk}(ABSENT)_{ijk} + \\
& \pi_{7jk}(MATERIAL)_{ijk} + \pi_{8jk}(WPLACE)_{ijk} + \pi_{9jk}(HB)_{ijk} + \\
& \pi_{10jk}(MEALS)_{ijk} + e_{ijk}
\end{aligned}
$$

Level-2 model

$$
\begin{aligned}
\pi_{0jk} &= \beta_{00k} + \beta_{01k}(HB\_2)_{jk} + \beta_{02k}(ZMTEST)_{jk} + r_{0jk} \\
\pi_{1jk} &= \beta_{10k} + \beta_{11k}(MTSEX)_{jk} + r_{1jk} \\
\pi_{2jk} &= \beta_{20k} \\
&\vdots \\
\pi_{10jk} &= \beta_{100k}
\end{aligned}
$$

Level-3 model

$$\begin{aligned}
\beta_{00k} &= \gamma_{000} + \gamma_{001}(PTRATIO)_k + \gamma_{002}(PBEHAVE)_k + \gamma_{003}(ABSENT_3)_k + u_{00k} \\
\beta_{01k} &= \gamma_{010} \\
\beta_{02k} &= \gamma_{020} \\
\beta_{10k} &= \gamma_{100} \\
&\vdots \\
\beta_{60k} &= \gamma_{600} + u_{60k} \\
\beta_{70k} &= \gamma_{700} + u_{70k} \\
\beta_{80k} &= \gamma_{800} \\
\beta_{90k} &= \gamma_{900} + u_{90k} \\
\beta_{100k} &= \gamma_{1000} + u_{100k}
\end{aligned}$$

Equation 5

Reading

Level-1 model

$$\begin{aligned}
[ZRALOCP]_{ijk} = \pi_{0jk} &+ \pi_{1jk}(AGE)_{ijk} + \pi_{2jk}(ENGLISH)_{ijk} + \\
&\pi_{3jk}(REPEAT)_{ijk} + \pi_{4jk}(RHWORKC)_{ijk} + \pi_{5jk}(ABSENT)_{ijk} + \\
&\pi_{6jk}(MATERIAL)_{ijk} + \pi_{7jk}(WPLACE)_{ijk} + \pi_{8jk}(HB)_{ijk} + \\
&\pi_{9jk}(MEALS)_{ijk} + e_{ijk}
\end{aligned}$$

Level-2 model

$$\begin{aligned}
\pi_{0jk} &= \beta_{00k} + \beta_{01k}(PABSEN_2)_{jk} + \beta_{02k}(ZRALOCT)_{jk} + r_{0jk} \\
\pi_{1jk} &= \beta_{10k} + r_{1jk} \\
\pi_{2jk} &= \beta_{20k} \\
\pi_{3jk} &= \beta_{30k} \\
\pi_{4jk} &= \beta_{40k} + r_{4jk} \\
\pi_{5jk} &= \beta_{50k} \\
&\vdots \\
\pi_{9jk} &= \beta_{90k}
\end{aligned}$$

Level-3 model

$$\beta_{00k} = \gamma_{000} + \gamma_{001}(PTRATIO)_k + \gamma_{002}(PBEHAVE)_k + u_{00k}$$
$$\beta_{01k} = \gamma_{010} + u_{01k}$$
$$\beta_{02k} = \gamma_{020}$$
$$\beta_{10k} = \gamma_{100}$$
$$\beta_{11k} = \gamma_{110}$$
$$\beta_{20k} = \gamma_{200} + u_{20k}$$
$$\beta_{30k} = \gamma_{300} + u_{30k}$$
$$\beta_{40k} = \gamma_{400}$$
$$\beta_{50k} = \gamma_{500} + \gamma_{501}(RHWORK\_3)_k + u_{50k}$$
$$\beta_{60k} = \gamma_{600} + u_{60k}$$
$$\beta_{70k} = \gamma_{700}$$
$$\beta_{80k} = \gamma_{800} + u_{80k}$$
$$\beta_{90k} = \gamma_{900} + u_{90k}$$

Equation 6

5.2 Estimating school effects

In this study, the school effects were estimated using a ‘subtraction’ method (see Raudenbush and Willms 1995; pp.318-9). For Type A effects, this involved controlling only for pupil characteristics, leaving school context, policy and practice unspecified. Basically, this involved running the models described in Equations 5 and 6 (for mathematics and reading respectively) with level-1 predictors but excluding the variable ‘Homework corrected’ because homework was considered part of school policy and practice. No predictors were entered in any other levels of the hierarchy.

Type B effects were estimated by controlling for pupil characteristics and school context, leaving school policy and practice unspecified. For mathematics, Type B effects were estimated using the model described in Equation 5 above but excluding the variables ‘Homework corrected’ (at level-1) and Mathematics tests (at level-2) because these variables were considered part of school policy and practice. Similarly, Type B effects for reading were estimated using the model described in Equation 6 but excluding variables that were part of school policy and practice. Consequently, the variable ‘Homework corrected’ (at level-1) was excluded in the estimation of Type B effects for reading. For the same reason, the interaction effect involving ‘Days absent’ and ‘Average Homework corrected’ (see Equation 6) was dropped in the estimation of Type B effects for reading.

In the HLM runs used to estimate school effects, all pupil-level predictor variables were grand-mean-centred in the HLM analyses so that the intercept terms would represent the mean mathematics and reading scores for schools (Willms and Raudenbush, 1989; Kreft, De Leeuw, Aiken, 1995). In addition, the empirical Bayes (EB) residuals from the HLM runs used to estimate school effects were exported into the AM computer program (AIR and Cohen, 2003) for further analyses to compare school effects across provinces in Kenya.

There are at least two problems associated with the above procedures used to estimate school effects.

First, Raudenbush and Willms (1995) noted that the estimation of school effects under the subtraction method could only be achieved without bias if 1) pupils were randomly assigned to schools and schools were randomly assigned to context conditions, and 2) if the school context and policies were orthogonal. Raudenbush and his colleague argued that a preferred alternative procedure would be to estimate school effects using ‘addition’ method. In order to estimate school effects using the addition method, all the relevant variables (that is, pupil background characteristics, school context, policy and practice) should be identified and accurately measured so that the model given by Equation 1 was fully specified. For this study, all the relevant data for school policy and practices were not available, and therefore, the subtraction method was used to estimate school effects.

Second, estimation of school effects as described above ignore the fact that schooling is longitudinal because the model used only captures school situation at one point in time. A preferred model may involve data collected from the same schools on several occasions. Such a model allows the performance of schools over time and the reliability of the school effects to be examined. Data needed for development of such a model only becomes available with subsequent SACMEQ studies.

6 Results and discussion

In the next two sub-sections, results of the analyses described above are presented and interpreted. The first sub-section focuses on the results of the fixed effects (path or regression coefficients) for each variable included in the final model while the second sub-section focuses on the results of variance partitioning and variance explained.

6.1 Variables in the final model

Estimates of fixed effects of the variables included in the final three-level models for mathematics and reading are given in *Table 2*. In interpreting the results in *Table 2*, the following three points are worth noting. First, all the coefficients displayed in *Table 2* are significant at $p \leq 0.05$ because their values taken in absolute terms are more than twice their standard errors. The standard errors given are those of the metric coefficients. Because weighting (using sampling weights) was undertaken in the construction of the SSM files, the HLM program took into consideration the design of this study in the computation of these standard errors.

<Insert Table 2 about here>

Second, the signs of metric and standardized coefficients indicate directions of effects and can be interpreted meaningfully if the coding of the variable is considered. For example, negative coefficients for AGE (Age in months) indicate that younger pupils were estimated to achieve better in mathematics and reading than older pupils. The positive coefficients for

MEALS (Meals per week) indicate that pupils who ate more meals per week were estimated to achieve better in mathematics and reading than pupils who ate fewer meals per week.

Third, absolute values of standardized coefficients (called 'effect size') can be used to rank variables in terms of their relative degree of influence on the outcome within the same sample, while those of metric coefficients can be used to compare different samples with each other (see Hox, 1995, p.26). Generally, in research studies in education, a standardized regression coefficient is considered important if its magnitude taken in absolute terms is ≥0.10. Thus, based on the standardized coefficients given in *Table 2*, it would appear that the key predictors of mathematics achievement among Standard 6 pupils in Kenya were Age in months (-0.16), Pupil's sex (-0.14), Pupil-teacher ratio (-0.14), Pupils' behaviour problems (0.14) and Average home background at the class-level (0.12). For reading, the key predictors were Pupil-teacher ratio (-0.19), Age in months (-0.18) and Home background at pupil-level (0.10).

In the following sub-sections, summaries of the effects recorded in *Table 2* on achievement in mathematics and reading among Standard 6 pupils in Kenya at the various levels of hierarchy are discussed. In these sub-sections, it is assumed that pupils differed only in the factor being considered and all other factors were equal. The numbers given in parenthesis in these sub-sections are the metric coefficients and their standard errors.

6.1.1 Pupil-level model

At the pupil level, it can be seen from the results given in *Table 2* that mathematics achievement was directly influenced by 10 of the 15 pupil-level variables examined in this study. These 10 variables were Age in months, Pupil's sex, Speaking English, Grade repetition, Homework corrected, Days absent, Material, Working place, Home background and Meals per week. All but one (Pupil's sex) of these 10 pupil-level variables also had significant influences on achievement in reading.

In summary, the following effects on achievement in reading and mathematics were recorded among Kenyan Standard 6 pupils when other factors were equal.

Age in months: Younger pupils were estimated to achieve better in mathematics (-0.78, 0.09) and reading (-0.88, 0.07) than their older counterparts. Being older in Standard 6 was therefore a distinct disadvantage. It was possible that this effect might have been a consequence of grade repetition by the less able pupils. However, this was unlikely because grade repetition was controlled for in the analysis after being found to be statistically significant (see below). Thus, parents should ensure that all children enter school at the right age. In addition, education policy should emphasize that children should enter school at the designed age.

Pupil sex: Boys were estimated to achieve better than girls in mathematics (-25.55, 2.92) but not in reading. This problem should be examined in depth by the Quality Assurance and Standards Division and the primary school mathematics specialists. In addition, the Ministry of Education, Science and Technology (MoEST) should commission studies to examine the reasons for the poor performance of girls in mathematics and to identify ways of correcting this problem.

Speaking English: Pupils who always spoke English (the language of the test) outside school were estimated to achieve better in mathematics (8.89, 4.44) and reading (19.59, 4.20) than pupils who never spoke English outside school. Clearly, it helped a great deal if the pupils spoke the language of the school (English) at home especially in their reading performance. Schools should therefore encourage this but at the same time the schools should maintain pupils' interest in Swahili, the other national language.

Grade repetition: Pupils who had never repeated a grade were estimated to achieve better in mathematics (-9.01, 2.65) and reading (-17.06, 2.60) than pupils who had repeated a grade one or more times. This relationship is of interest because it was observed even after the influence of the variable Age in months (see above), had been controlled for in the models. It is an open secret that some school heads (especially those in private schools and some high performing public schools) have grade repetition policies to ensure high performance of their schools in the Kenya Certificate of Primary Education (KCPE) examination. Therefore, the MoEST in collaboration with Kenya National Examination Council (KNEC) should find ways of discouraging grade repetition as a method of uplifting school performance in the KCPE examination. In other words, education policy should discourage grade repetition, without lowering standards of achievement.

Homework corrected: Pupils who were given homework more frequently and had it corrected were estimated to achieve better in mathematics (8.06, 2.47) and reading (7.94, 2.44) than pupils who were given homework and had it corrected less frequently. Therefore, education policy should specify the amount of homework that teachers should give to pupils per week. In addition, the MoEST, through school heads and Quality Assurance and Standards Division, should make sure that all teachers corrected the homework given to pupils.

Days absent: Pupils who were never (or rarely) absent from school were estimated to achieve better in mathematics (-1.67, 0.45) and reading (-2.13, 0.50) than those pupils who were frequently absent from school. Obviously, pupils who were regular absentees received fewer hours of instruction and therefore were highly likely to achieve at a lower level compared to the rest of their classmates.

Policy should concentrate on low absenteeism in schools. In addition, further analyses should be carried out on these data to examine the relationship between absenteeism and other factors among Standard pupils in Kenya with the aim of equipping teachers, parents and policy-makers with research information that would assist them curb this problem.

Material: Pupils who had most learning materials (pencils, pens, exercise books, notebooks, erasers and rulers) were estimated to achieve better in mathematics (-7.08, 1.44) and reading (-8.29, 1.43) than pupils who had hardly any learning materials. Clearly, it is important for pupils to have these basic learning materials for improved achievement in reading and mathematics as well as for academic progress in general. Under the Free Primary Education (FPE) program in Kenya, the government now provides these learning materials to pupils, which is a major step towards solving this problem. Before the introduction of FPE program in 2003 in Kenya, provision of these learning materials was left to parents.

Working place: Pupils who had their own working places in class (for sitting and writing) were estimated to achieve better in mathematics (5.41, 0.98) and reading (4.71, 0.92) than pupils who shared working places or had no working places in class. This implies that the MoEST should ensure that every class has sufficient working places. Surely, pupils are less motivated to learn if they have to spend the whole day in uncomfortable sitting and writing places because of lack of furniture or over crowding in classrooms.

Home background: Pupils from homes with better quality houses, many possessions (wealthy) and more educated parents were estimated to achieve better in mathematics (6.70, 2.34) and reading (9.60, 1.97) than pupils from homes with low quality houses, few or no possessions (poor) and less educated parents. Education policy should focus on finding ways to encourage low socioeconomic status parents to take an interest in their children's schoolwork. Hence, the government should initiate long-term programs (e.g. attractive adult literacy classes, home electrification projects and small-scale economic projects) aimed at eradicating poverty in order to raise levels of educational achievement over time.

Meals per week: Pupils who ate more meals per week were estimated to achieve better in mathematics (1.47, 0.43) and reading (1.29, 0.40) than pupils who ate fewer meals per week. Parents should ensure children get enough meals per week so that the children have adequate energy for learning. In addition, the government should assist parents by starting School Feeding Programs (SFP) to ensure that all children receive enough meals per week so that they can learn effectively. It is suspected that effective SFP could lower pupil absenteeism problems in schools and, to some degree, could also lower pupils' behaviour problems in schools.

For mathematics, there was a significant interaction effect involving Age in months (a pupil-level variable) and Teacher sex (a class-level variable). For reading, there was a significant interaction effect involving Days absent (a pupil-level variable) and Average homework corrected (a school-level variable). These interaction effects are discussed in the paragraphs that follow. Graphical representations of these interaction effects are given in *Figure 1* and *Figure 2* to enhance discussion. The coordinates of the graphs shown in *Figure 1* and *Figure 2* were calculated from the final estimation of the fixed effects obtained from the final models (results given in *Table 2*). Lietz (1996) has described the procedure employed to calculate the coordinates of such graphs. A book by Aiken and West (1996) was consulted for the interpretation of this interaction effect.

The graphical representation in *Figure 1* shows that young pupils were generally likely to achieve better in mathematics than their older counterparts regardless of whether they were taught by male or female teachers. More importantly, younger pupils who were taught by female teachers were likely to achieve better than younger pupils who were taught by male teachers. On the contrary, older pupils were likely to achieve better in mathematics if taught by male teachers than their age mates who were taught by female teachers. It is possible that this complex relation was due in part, at least, to problems of discipline in the classrooms with older boys.

The graphical representation in *Figure 2* shows that pupils who were rarely absent from school had much better reading achievement if they were in schools where homework was given more frequently and had the homework given corrected than if they were in schools where homework was rarely given and rarely corrected. However, for frequent absentees, *Figure 2* shows that there was no difference in reading achievement regardless of the amount of homework and homework corrections in schools. Obviously, to benefit from homework given it is essential for pupils to attend school more regularly.

<*Insert Figure 1 about here*>

<*Insert Figure 2 about here*>

6.1.2 Class-level model

At the class-level, out of the 25 class-level variables examined in the HLM analysis (listed in *Table 1*), two had significant influences on achievement in mathematics (i.e. Average home background status and Mathematics tests) and two on reading achievement (Average days absent and Teacher score).

Thus, other factors being equal, the following effects were identified (*Table 2*) regarding the achievement in mathematics and reading of Standard 6 pupils in Kenya.

Average home background: Pupils in classes with a majority of the pupils from homes with good quality houses, many possessions and more educated parents were estimated to perform better in mathematics (12.45, 4.73) than pupils in classes with a majority of the pupils from homes with poor quality houses, fewer possessions and less educated parents.

Mathematics tests: Pupils in classes where mathematics tests were given more frequently were likely to perform better in mathematics (6.17, 3.05) compared to their counterparts in classes where mathematics tests were given less frequently. This effect of mathematics tests on pupil achievement should be interpreted with caution because it could be misused to pressurize teachers and schools to adopt a frequent testing policy, which might not necessarily be for the better. A frequent testing policy has the potential of causing teachers to alter the content of their classroom instruction to match the tests. In other words, teachers could find themselves in a situation where they were compelled to focus on test-specific materials or coach pupils to be so-called ‘test-wise’ in KCPE subjects everyday and neglect other subjects and extra curricula activities that were equally important in the well being of the pupils. What is needed is a clear policy from the MoEST on frequency of testing in schools.

Average days absent: Pupils in classes in which a majority of the pupils were never (or rarely) absent from school were estimated to perform better in reading (-4.32, 1.86) than pupils in classes in which a majority of the pupils were more often absent from school. This class-level relationship is of interest because it was observed even after the influence of the variable Days absent (see *Table 2*) was controlled for in the model at the pupil-level. This implied that a high rate of absenteeism at the class-level affected regular attendees within the class as well.

Teacher score: Pupils taught by teachers who had higher reading scores were likely to achieve better in reading (0.12, 0.05) than pupils taught by teachers with lower reading scores. Thus, a teacher-training and recruitment policy should be put in place to ensure that all teachers have excellent subject matter knowledge. Such a policy involves providing teaching permits (renewable periodically) for those who meet the desired levels of professional training and subject matter knowledge. For serving teachers, the MoEST should establish in-service training programs for teachers in both public and private schools to be undertaken during the school holidays. These in-service training programs could form the basis not only for promotion of teachers but also for the renewal of teaching permits.

6.1.3 School-level model

The results in *Table 2* show that three of the school-level variables examined in the multilevel analyses (listed in *Table 1*) had a significant influence on achievement in mathematics. These three school-level variables were Pupil-teacher ratio, Pupils' behaviour problems, and Average days absent. Apart from Average days absent, the other two school-level variables that had significant influences on achievement in mathematics also had significant influences on achievement in reading.

Thus, other factors being equal, the following effects were identified (*Table 2*) regarding the achievement in mathematics and reading of Standard 6 pupils in Kenya.

Pupil-teacher ratio: Pupils in schools with smaller pupil-teacher ratios were estimated to perform better in mathematics (-1.37, 0.43) and reading (-1.89, 0.35) than pupils in schools with larger pupil-teacher ratios. This implies that that the MoEST should employ more teachers to lower pupil-teacher ratio. In addition, the MoEST in collaboration with the Teachers' Service Commission should ensure equity in pupil-teacher ratio among schools. An examination of the pupil-teacher ratio is an urgent matter because the introduction of Free Primary Education in 2003 resulted in substantial influx of pupils in schools in some parts of the country and this may have an adverse effect on the quality of education provided in affected schools.

Pupils' behaviour problems: Pupils in schools with little or no pupils' behaviour problems were estimated to perform better in mathematics (12.54, 2.72) and reading (8.24, 3.27) than their counterparts in schools with many pupils' behaviour problems. Education policy should concentrate on minimizing pupil behaviour problems in schools.

Average days absent: Pupils in schools in which a majority of the pupils were never (or rarely) absent from school were estimated to perform better in mathematics (-5.74, 2.35) than pupils in schools in which a majority of the pupils were more often absent from school.

6.2 Variance partitioning and variance explained

The variance components from the final and null models are presented in *Table 3* in rows ‘a’ and ‘b’ respectively. From the information in *Table 3* rows ‘a’ and ‘b’, the information presented in rows ‘c’ to ‘f’ were calculated. A discussion of the calculations involved here is to be found in Raudenbush and Bryk (2002, pp.69-95).

Thus, results in *Table 3* show that the percentages of variances available at the pupil, class and school levels were 61.1, 5.1 and 33.8 for mathematics respectively, 52.9, 3.0 and 44.0 and reading for respectively. These percentages of variance of pupil scores at the various levels of the hierarchy were the maximum amounts of variance available at those levels that could be explained in subsequent analyses.

<Insert Table 3 about here>

The result of variance partitioning show that there were only small differences (5 per cent or less) between classes within schools in term of their pupils’ achievement in mathematics and reading in Kenya. In addition, these results show that the variance between schools in Kenya was comparable to what is generally reported at similar grade levels in other developing countries. For example, Willms and Somers (2001), utilizing data from grades 3 and 5 pupils from 13 Latin American countries, found that the variance between schools in mathematics achievement ranged from 19.5 to 41.2 per cent.

At the school-level, the variance of pupil’s scores for mathematics (33.8 per cent) was noticeably smaller than the variance of pupil’s scores for reading (44.0 per cent). This could be in part due to the fact that, unlike mathematics skills, the reading skills are not only learned in school (see Hungi, 2003). The pupils could acquire some of the reading skills outside their schoolwork (for example, at home watching television, reading advertisement and mails), and the opportunities to acquire reading skills outside school differ for schools in different locations.

The predictors included in the final model for mathematics explained 22.1 per cent of 61.1 per cent of the variance available at the pupil-level and that was equal to 13.5 per cent (that is, 22.1 × 61.1) of the total variance explained at the pupil-level. Similarly, for the same mathematics model, predictors included in the final model explained 1.7 per cent (that is, 34.3 per cent of 5.1 per cent) at the class-level, and 18.9 per cent (that is, 56.0 per cent of 33.8 per cent) at the school-level. Thus, the total variance explained by the predictors included in the final model for mathematics was $13.5 + 1.7 + 18.9 = 34.2$ per cent, which left 65.6 per cent of the total variance unexplained. Likewise, the percentages of variance explained by the predictors included in the final model for reading at the pupil, class and school levels were 11.9, 2.2 and 27.0 respectively. Thus, the predictors included in the final model for reading explained 41.0 per cent of the total variance, which left 59.0 per cent of the total variance unexplained.

In addition, the results in *Table 3* (row 'f') show that, based on the multilevel models developed in this study, the amounts of variances left unexplained were large (≥59 per cent) for both mathematics and reading. Clearly, there is a need for a further study to examine what other important factors were left out of this study. Such factors would assist in development models that are the most appropriate for explaining pupil achievement in Kenya and which maximize the total variance explained.

Two key factors that were not examined in this study were prior achievement and pupil mobility (transience). Of these two factors, prior achievement is more vital because nearly all studies on factors influencing student achievement show that prior achievement is highly correlated with later student achievement, with students with high prior achievement scores achieving higher scores in subsequent achievement tests. In addition, most studies report prior achievement as the highest contributing factor in the prediction of student achievement (e.g. Ethington, 1992; Reynolds and Walberg, 1992; Gill and Reynolds, 1999; Fuchs *et al*. 2000).

However, some debate exists regarding the appropriateness of using a prior achievement variable in explaining variance in pupil achievement. This debate arises when it is thought to be difficult to obtain reliable prior achievement information especially for studies conducted at the points of entry to primary or secondary schools (Teddlie, Reynolds and Sammons, 2001). In Kenya, however, some schools conduct comprehensive tests (referred to as ‘interviews’) for selecting Standard 1 pupils and also for selecting pupils wishing to transfer to the school at other grade levels. This practice prevails mostly in prestigious primary schools (both private and public) especially those that have pupil intake at middle grade levels (i.e. Standards 4 – 6). This practice is also common in schools with good history of performance in the KCPE.

Under the above circumstances, where some schools get top pupils from the less advantaged schools, it is important that those wishing to explain variance in pupil achievement in Kenya should consider a measure of prior achievement (e.g. pupils’ results obtained from school entry interviews).

In addition, prior achievement could be critical to those wishing to compare performance of schools in Kenya because, within the context of value added, the progress made by pupils who move between schools would not be due to the efforts of one school alone. Thus, adjustments should be made to cater for schools that have their pupil intake at middle grade levels.

Having said the above regarding prior achievement, it is essential for the MoEST to provide clear directions regarding the tests conducted by schools for purposes of selecting pupils for intake. This is because, more often than not, schools charge exorbitant fees for these tests and these tests are major sources of corruption in public schools. Moreover, it is unclear what should happen to a child who fails these tests and is therefore rejected by schools as early as Standard 1.

6.3 School effects across provinces

Figure 3 and *Figure 4* show the histogram plots of the school effects by Province for mathematics and reading respectively. The ‘Raw effects’ in the histogram plots were obtained from HLM analyses of the null models. In other words, the Raw effect is school performance before adjustments were made for any factors. On the other hand Type A effect is school performance after adjustments were made to cater for the contributions of pupil-background factors while Type B effect is school performance after adjustments were made to cater for the contributions of pupil-background, class-context and school-context factors.

<Insert Figure 3 about here>

<Insert Figure 4 about here>

In the estimation of the standard errors shown in the histograms plots (*Figure 3* and *Figure 4*), multilevel nature of the school effect data (i.e. schools nested within province) were taken into consideration using AM (AIR and Cohen, 2003) computer software.

When interpreting the results displayed in *Figure 3* and *Figure 4,* it is important to consider the following two issues. First, in general, zero is the average of the school effects and schools with values with positive signs are considered to be relatively effective when compared to the average, while schools with values with negative signs are considered to be relatively ineffective when compared to the average. That is, schools with positive values are likely to contribute more to the increase in pupil achievement, while schools with negative values tend to contribute less to the increase in pupil achievement. Importantly, these comparisons between schools are relative ones. Thus, the use of the descriptive terms ‘effective’ or ‘ineffective’ can be misleading because, based on some absolute criterion, all the schools being compared can be performing poorly or all the schools can be performing well (Coe and Fitz-Gibbon, 1998).

Second, the error bars shown in *Figures 3* and *4* are twice the standard errors of the school effects. Therefore, school effects were considered to be significantly different from average at $p \leq 0.05$ level if their error bars did not cross the horizontal zero line in these diagrams. Similarly, across provinces, school effects were considered significantly different from each other if their error bars did not overlap.

Thus, for mathematics, *Figure 3* shows that, if performances of schools were compared using raw scores across the eight provinces, it could be concluded that schools in Nairobi performed above average and that schools in Western performed below average. However, to those interested in the value added by schools, such conclusions could be misleading because, after removing the effects of pupil characteristics (Type A) and also after removing the effects of pupil characteristics and school context (Type B), *Figure 3* provides no evidence of possibilities of significant differences between performances of schools in the eight provinces based on mathematics achievement of their pupils in Standard 6.

For reading, however, *Figure 4* shows that schools in Nairobi performed above average even after adjustments were made (Types A and B). Interestingly, with exceptions of schools in Coast and Eastern, *Figure 4* provides evidence that there were significant differences between performances of schools in Nairobi and performances of schools in the other provinces based on Types A and B effects for reading.

There is the need for further studies to investigate why schools in Nairobi performed better in reading compared to schools in the other provinces because such studies could yield information that could be useful in improving reading achievement across Kenya. One possible explanation could lie in the fact that, unlike the other provinces in Kenya, Nairobi is purely metropolitan and has extensive facilities (such as a wide variety of television stations, billboards and advertisement mails), which can help pupils to acquire some reading skills outside their schoolwork. It should be remembered that, unlike mathematics skills, reading skills are not purely learned in school and pupils could acquire some reading skills outside their schoolwork.

Nevertheless, the results of analyses represented in this section should be of interest to those wishing to examine performance of schools in Kenya based on pupils test scores. Clearly, it would be misleading to compare schools in Kenya using average scores without adjusting the scores for the differences between schools in their pupil intake.

7 Conclusions

The purposes of this study were to 1) identify the key pupil-, class- and school-related factors that influence achievement in mathematics and reading among Standard 6 pupils in Kenya, 2) develop multilevel models that could be used to explain some of the variance in mathematics and reading achievement among Standard 6 pupils in Kenya, 3) examine school effects across provinces in Kenya based on the multilevel models developed in this study.

In order to achieve the above purposes, a three-level model was hypothesized and analyzed using HLM5 software for each of the two outcome measures (mathematics and reading). In addition, the performances of schools across provinces in Kenya were examined based on the three-level model and using a value-added approach.

Based on the magnitudes of effect sizes of the variables in the final mathematics model, it was found that the key predictors of mathematics achievement among Standard 6 pupils in Kenya were Age in months (-0.16), Pupil's sex (-0.14), Pupil-teacher ratio (-0.14), Pupils' behaviour problems (0.14) and Average home background at the class-level (0.12). For reading, the key predictors were Pupil-teacher ratio (-0.19), Age in months (-0.18) and Home background at individual-level (0.10).

The results of variance partitioning showed that the percentages of variances available at the pupil, class and school levels were 61.1, 5.1 and 33.8 for mathematics respectively, and 52.9, 3.0 and 44.0 for reading respectively. These results of variance partitioning showed that, at Standard 6 primary school level in Kenya, there were only small differences between classes within school but considerable differences between schools in terms of mathematics and reading achievement. In addition, at the school-level, these results showed that the variance between schools in Kenya was comparable to what is generally reported at similar grade levels in other developing countries.

For both mathematics and reading, the multilevel models developed in this study explained only small percentages of the total variances available (34.2 and 41.0 for mathematics and reading respectively). The large percentages of variance left unexplained (65.8 and 59.0 for mathematics and reading respectively) strongly indicated that there were other important factors that influenced pupils' achievement in mathematics and reading (such as prior achievement and pupil mobility), which were not included in the models developed in this study. Consequently, it was suggested that, future studies wishing to explain differences in pupil achievement in Kenya, should consider including prior achievement variable into the model.

Apart from examining what other important factors were left out in this study, it is important to consider that some of the between-school variance may be attributed to the differences between the provinces that the schools belonged to. In other words, the inclusion of a province-level in the analyses can help to explain some of the school-level variance. Consequently, it would be profitable to repeat these analyses based on a four-level model (pupil, class, school and province). Such analyses can be carried out using MLwiN (Browne, Healy, Cameron, and Charlton, 2001) software because the current versions of HLM program cannot handle more than three levels of analyses.

Finally, results showed that, based on raw effects (unadjusted average test scores of pupils), there were significant differences in the performances of schools across provinces in Kenya, with schools in Nairobi performing above average and schools in Western performing below average. However, with an exception of Nairobi (but only in reading), the results showed that, after removing the effects of pupil characteristics (Type A) and also after removing the effects of pupil characteristics and school context (Type B), there were no significant differences between performances of schools across provinces in Kenya. It was suggested that the metropolitan nature of Nairobi might have assisted pupils to acquire some reading skills outside their schoolwork.

8 References

Aiken, L.S.; West, S.G. 1996. *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage Publications.

AIR; Cohen, J. 2003. *AM* (Version 0.5.13). Washington, DC: American Institutes for Research.

Browne, W.; Healy, M.; Cameron, B.; Charlton, C. 2001. *MLwiN* (Version 1.10.0007). London: University of London, Institute of Education.

Bryk, A.S.; Raudenbush, S.W. 1992. *Hierarchical linear models: Applications and data analysis methods*. Newbury Park: Sage Publications.

Coe, R.; Fitz-Gibbon, C.T. 1998. School effectiveness research: Criticisms and recommendations. *Oxford Review of Education*, *24* (4), 421–438.

Comber, L.C.; Keeves, J.P. 1973. *Science education in nineteen countries: An empirical study*. Stockholm: Amquist and Wiksell.

Ethington, C.A. 1992. Gender differences in a psychological model of mathematics achievement. *Journal for Research in Mathematics Education*, *23* (2), 166–181.

Fuchs, L.S.; Fuchs, D.; Karns, K.; Hamlett, C.L.; Dutka, S.; Katzaroff, M. 2000. The importance of providing background information on the structure and scoring of performance assessments. *Applied Measurement in Education*, *13* (1), 1–34.

Gill, S.; Reynolds, A.J. 1999. Educational expectation and school achievement of urban African-American children. *Journal of School Psychology*, *37* (4), 403–424.

Howie, S. 2002. *English language proficiency and contextual factors influencing mathematics achievement of secondary school pupils in South Africa*. Enschede: Print Partners Ipskamp.

Hox, J.J. 1995. *Applied multilevel analysis*. Amsterdam: TT-Publikaties.

Hungi, N. 2003. *Measuring school effects across grades*. Adelaide: Shannon Research Press.

Husén, T. 1967. *International study of achievement in mathematics* (Vol.2). Stockholm: Almqvist and Wiksell.

Keeves, J.P. 1975. The home, the school and achievement in mathematics and science. *Science Education*, *59* (4), 207–218.

Kreft, I.G.G.; De Leeuw, J.; Aiken, L.S. 1995. The effects of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research, 30* (1), 1–21.

Lietz, P. 1996. *Changes in reading comprehension across cultures and over time*. Minister: Waxman.

McPherson, A. 1993. *Measuring added value in schools* (pp.1-4). Edinburgh, UK: National Commission on Education.

Postlethwaite, T.N.; Wiley, D.E. 1991. *The IEA study of Science II. Science achievement in twenty-three countries*. Oxford: Pergamon.

Raudenbush, S.W.; Willms, J.D. 1995. The estimation of school effects. *Journal of Educational and Behavioral Statistics, 20* (4), 307–335.

Raudenbush, S.W.; Bryk, A.S. 1994. Hierarchical linear models. In T. Husén and T.N. Postlethwaite (Eds.), *International Encyclopedia of Education: Research and Studies* (2nd ed., pp. 2590–2596). Oxford: Pergamon.

Raudenbush, S.W.; Bryk, A.S. 2002. *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Raudenbush, S.W.; Bryk, A.S.; Cheong, Y.F.; Congdon, R.T. 2000*a*. *HLM5: Hierarchical linear and nonlinear modeling* (Version 5.01.2067.1). Lincolnwood, IL: Scientific Software International.

Raudenbush, S.W.; Bryk, A.S.; Cheong, Y.F.; Congdon, R.T. 2000*b*. *HLM5: Hierarchical linear and non-linear modeling (User guide)*. Lincolnwood, IL: Scientific Software International.

Reynolds, A.J.; Walberg, H.J. 1992. A process model of mathematics achievement and attitude. *Journal for Research in Mathematics Education, 23* (4), 306–328.

Teddlie, C.; Reynolds, D.; Sammons, P. 2001. The methodology and scientific properties of school effectiveness research. In C. Teddlie and D. Reynolds (Eds.), *The International Handbook of School Effectiveness Research* (pp. 55–133). London: Routledge Falmer.

Willms, D.J.; Raudenbush, S.W. 1989. A longitudinal hierarchical linear model for estimating school effects and their stability. *Journal of Educational Measurement, 26*, 209–232.

Willms, D.J.; Somers, M.A. 2001. Family, classroom, and school effects on children's educational outcomes in Latin America. *School Effectiveness and School Improvement, 12* (4), 409–445.

Yang, M.; Goldstein, H.; Rath, T.; Hill, N. 1999. The use of assessment data for school improvement purposes. *Oxford Review of Education, 25* (4), 469–483.

Table 1 Variables tested on each level of the hierarchy

Level	Variable of interest	Variables tested in HLM
Pupil	Pupil's sex (0=Boy; 1=Girl)	SEX
	Age in months	AGE
	Speaking English (0=Never; 1=Sometimes/always the time)	ENGLISH
	Books at home	BOOKS
	Meals per week	MEALS
	Grade repetition (0=Never repeated; 1=Repeated at least once)	REPEAT
	Being asked to calculate/Being asked to read (0=Never/sometimes; 1=Most of the time)	CALC, READ
	Being asked questions about subject (0=Never/sometimes; 1=Most of the time)	QUESTM, QUESTR[a]
	Extra tuition (0=No extra tuition; 1=Takes extra tuition on a subject)	EXTRAT
	Homework corrected (1=Never ; 2=Sometimes; 3=Most of the time/always)	MHWORKC, RHWORKC[a]
	Days absent (0=Lowest, . . . , 26=Highest)	ABSENT
	Textbook ownership (0=No textbook or share; 1=Own textbook)	TEXTM, TEXTR[a]
	Working place (Writing and Sitting)	WPLACE[b]

	Material (e.g. Exercise book, pencils, and eraser)	MATERIAL[b]
	Home background	HB[b]
Class	Proportion of girls	SEX_2
	Average age in months	AGE_2
	Average speaking English	ENGLIS_2
	Average books at home	BOOKS_2
	Average meals per week	MEALS_2
	Average grade repetition	REPEAT_2
	Average homework corrected	MHWORK_2, RHWORK_2[a]
	Average days absent	ABSENT_2
	Average textbook ownership	TEXTM_2, TEXTR_2[a]
	Average working place	WPLACE_2
	Average material	MATERI_2
	Average home background	HB_2
	Class size	MCSIZE, RCSIZE[a]
	Teacher's sex (0=Male; 1=Female)	MTSEX, RTSEX[a]
	Teacher's age level	MTAGELVL, RTAGELVL[a]
	Teacher's training	MTQPROF, RTQPROF[a]
	Classroom resources	MTCLRES8, RTCLRES8[a]
	Teacher's teaching hours per week	MTHRTEAC, RTHRTEAC[a]
	Teacher's frequency of meeting parents	MTMEET, RTMEET
	Teacher's inspector and adviser visits	MTINSTOT, RTINSTOT[a]
	Teacher's possessions	MTHPOS13, RTHPOS13[a]

	Teacher's source of light	MTLIGHT, RTLIGHT[a]
	Teacher's home condition	MTCONDLI, RTCONDLI[a]
	Teacher score	ZMALOCT, ZRALOCT[a]
	Mathematics/Reading tests (0=1 per term; 1=2 or 3 per month; 2=1+ per week)	MTEST, RTEST[a]
School	Average pupils' age	AGE_3
	Proportion of girls in the school	SEX_3
	Average speaking English	ENGLIS_3
	Average books at home	BOOKS_3
	Average meals per week	ZPREGM_3
	Average grade repetition	REPEAT_3
	Average homework corrected	MHWORK_3, RHWORK_3[a]
	Average days absent	ABSENT_3
	Average textbook ownership	TEXTM_3, TEXTR_3 [a]
	Average pupil home background	HB_3
	Proportion of female teachers	SEX_3
	School type (1=Government; 2=Private)	SCHTYPE
	Head teacher's sex (0=Male; 1=Female)	HSEX
	Head teacher's age level	HAGELVL
	Head teacher qualification	HQTT
	Head teacher teaching hours per week	HEADTCH
	School location (0=Isolated/rural; 1=Town/Urban)	LOCATION

Pupil-teacher ratio	PTRATIO
Prop. of teachers having tertiary academic education	TCHACA
School size	SSIZE
Pupils' behaviour problems	PBEHAVE[b]
Teachers' behaviour problems	TBEHAVE[b]
School resources	RESOURCE[b]
School community contribution	COMUNITY[b]
Average teachers' score	ZMALOC_3, ZRALOC_3[a]
Borrowing of books (0=Pupils can not borrow; 1=Pupil can borrow)	BORROW

Notes: [a] Variable listed first was included in the mathematics model while the second variable was included in the reading model.

[b] Composite variable (see Appendix for variables involved in construction of this variable).

Suffix '_2' is used to indicate a class-level aggregated variable.

Suffix '_3' is used to indicate a school-level aggregated variable.

Table 2 Final fixed effects for mathematics and reading

a) Mathematics

Level	Variable name	Variable included	Std'zed	Metric	SE
School	Intercept			566.53	3.20
	Pupil-teacher ratio	PTRATIO	-0.14	-1.37	0.43
	Pupils' behaviour problems	PBEHAVE	0.14	12.54	2.72
	Average days absent	ABSENT_3	-0.08	-5.74	2.35
Class	Average home background	HB_2	0.12	12.45	4.73
	Mathematics tests	MTEST	0.05	6.17	3.05
Pupil	Age in months	AGE	-0.16	-0.78	0.09
	interaction with Maths teacher's sex	MTSEX	0.00	-0.33	0.16
	Pupil's sex	SEX	-0.14	-25.55	2.92
	Speaking English	ENGLISH	0.03	8.89	4.44
	Grade repetition	REPEAT	-0.05	-9.01	2.65
	Maths homework corrected	MHWORKC	0.05	8.06	2.47
	Days absent	ABSENT	-0.05	-1.67	0.45
	Material	MATERIAL	-0.08	-7.08	1.44
	Working place	WPLACE	0.06	5.41	0.98
	Home background	HB	0.07	6.70	2.34
	Meals per week	MEALS	0.06	1.47	0.43

b) Reading

Level	Variable name	Variable included	Std'zed	Metric	SE
School	Intercept			548.72	3.27
	Pupil-teacher ratio	PTRATIO	-0.19	-1.89	0.35
	Pupils' behaviour problems	PBEHAVE	0.09	8.24	3.27
Class	Average days absent	ABSENT_2	-0.07	-4.32	1.86
	Teacher score	ZRALOCT	0.09	0.12	0.05
Pupil	Age in months	AGE	-0.18	-0.88	0.07
	Speaking English	ENGLISH	0.07	19.59	4.20
	Grade repetition	REPEAT	-0.09	-17.06	2.60
	Reading homework corrected	RHWORKC	0.05	7.94	2.44
	Days absent	ABSENT	-0.07	-2.13	0.50
	interaction with Homework corrected	RHWORK_3	-0.02	-5.06	1.70
	Material	MATERIAL	-0.09	-8.29	1.43
	Working place	WPLACE	0.05	4.71	0.92
	Home background	HB	0.10	9.60	1.97
	Meals per week	MEALS	0.05	1.29	0.40

Notes: Suffix _2 is used to indicate a class-level aggregated variable.

Suffix _3 is used to indicate school-level aggregated variable.

Metric is unstandardized regression (path) coefficient.

Std'zed is standardized regression (path) coefficient, also called effect size.

All coefficients are significant at $p<0.05$.

Standard errors (SE) given are those of the metric coefficients.

Table 3 Variance partitioning and variance explained

	Pupil (N=3,299)	Class (N=320)	School (N=185)	Total
Mathematics				
[a] Null model	4623.54	383.77	2561.88	7569.19
[b] Final model	3602.13	252.03	1128.46	
[c] Variance available	61.1%	5.1%	33.8%	
[d] Variance explained	22.1%	34.3%	56.0%	
[e] Total variance explained	13.5%	1.7%	18.9%	**34.2%**
[f] Variance left unexplained	47.6%	3.3%	14.9%	65.8%
Reading				
[a] Null model	4244.86	243.15	3533.23	8021.23
[b] Final model	3290.72	69.45	1369.78	
[c] Variance available	52.9%	3.0%	44.0%	
[d] Variance explained	22.5%	71.4%	61.2%	
[e] Total variance explained	11.9%	2.2%	27.0%	**41.0%**
[f] Variance left unexplained	41.0%	0.9%	17.1%	59.0%

Figure 1 **Impact of the interaction effect of pupil Age in months with Teacher sex on Mathematics achievement**

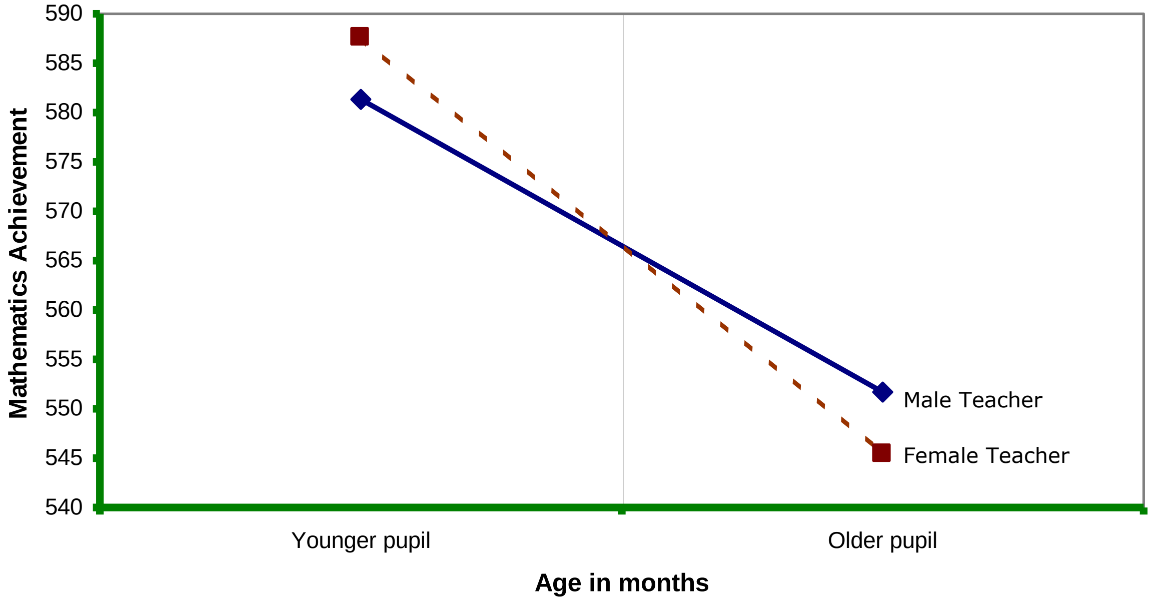


Figure 2 **Impact of the interaction effect of pupil's Days absent with Average homework corrected at the school level on Reading achievement**

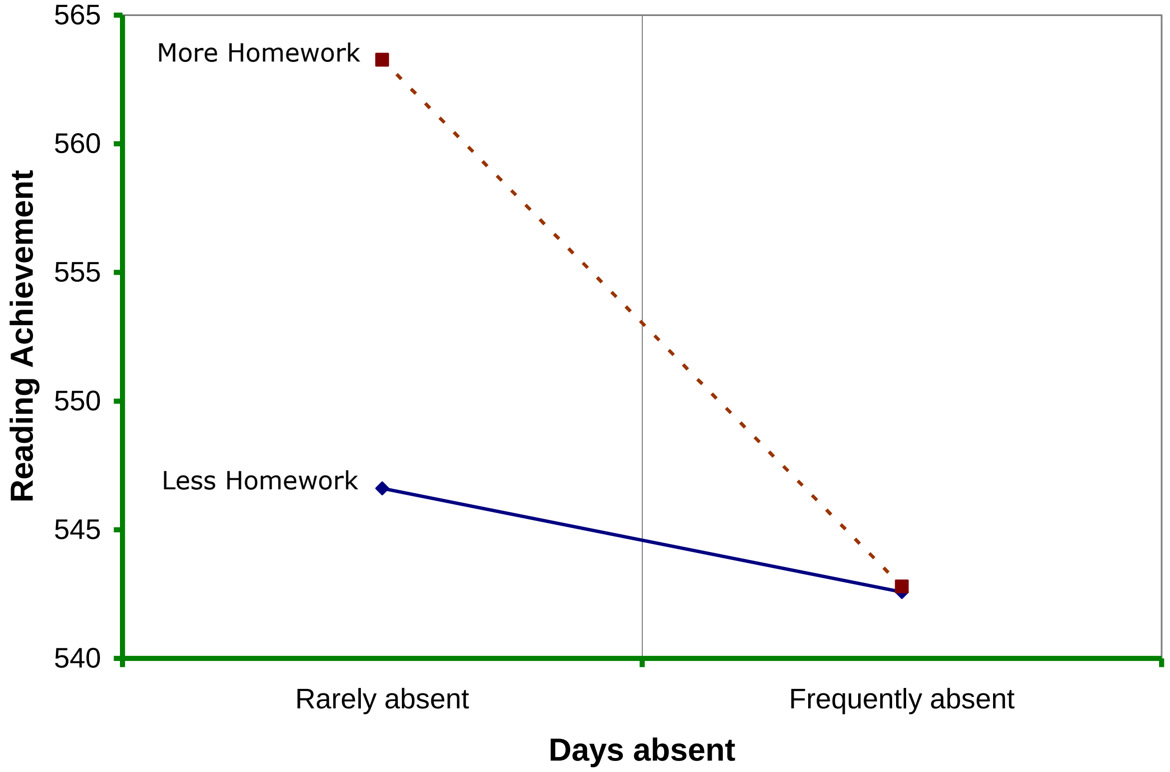
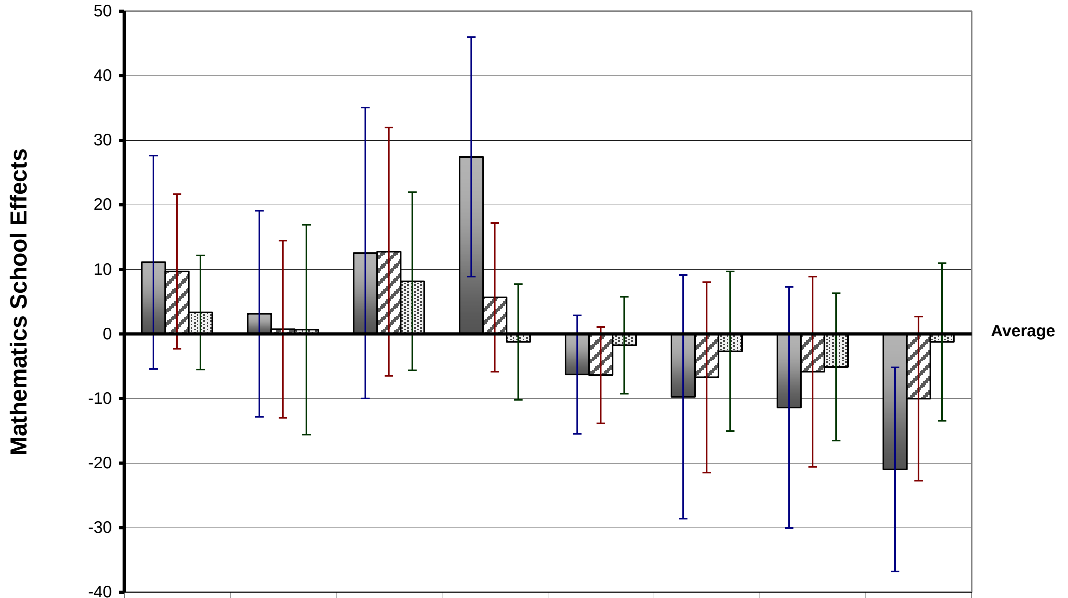
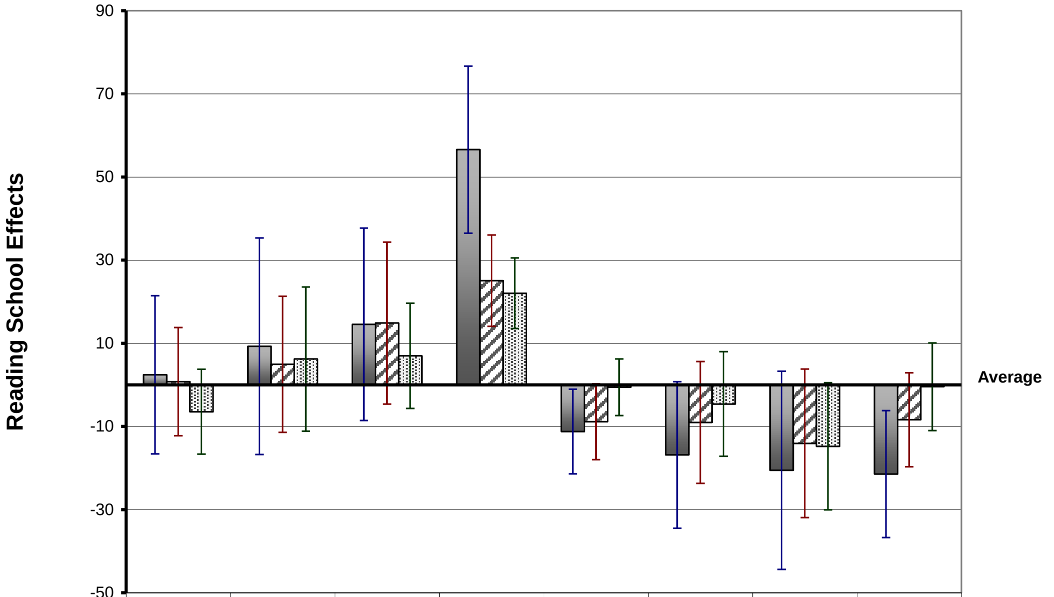


Figure 3 Mathematics school effects by Province



	Central	Coast	Eastern	Nairobi	N. Eastern	Nyanza	Rift Valley	Western
Raw effects	11.12	3.13	12.55	27.43	-6.29	-9.73	-11.38	-20.97
Type A effects	9.68	0.75	12.75	5.67	-6.38	-6.72	-5.84	-10.02
Type B effects	3.34	0.67	8.17	-1.23	-1.73	-2.68	-5.10	-1.23

Figure 4 Reading school effects by Province



	Central	Coast	Eastern	Nairobi	N. Eastern	Nyanza	Rift Valley	Western
Raw effects	2.43	9.29	14.57	56.59	-11.24	-16.86	-20.55	-21.45
Type A effects	0.78	4.95	14.87	25.07	-8.87	-9.05	-14.07	-8.40
Type B effects	-6.46	6.21	7.00	22.04	-0.58	-4.61	-14.78	-0.45

9 Appendix
Factor loadings of the variables used to construct composite variables

Factor	Variables	Loadings
HB		
	Possessions at home	0.80
	Quality of house	0.88
	Parents' education	0.63
	Pupil's source of light	0.84
WPLACE		
	Sitting place	0.75
	Writing place	0.75
PBEHAVE		
	Arrive late	0.62
	Skip class	0.56
	Dropout	0.41
	Classroom disturbance	0.57
	Cheating	0.67
	Language	0.76
	Vandalism	0.78
	Theft	0.69
	Bullying pupils	0.67
	Bullying staff	0.73
	Injure staff	0.60
	Sexually harass pupils	0.61
	Sexually harass teachers	0.44
	Drug abuse	0.59
	Alcohol abuse	0.54
	Fights	0.61
MATERIAL		
	No exercise books	0.62
	No notebooks	0.43
	No pencils	0.66
	No erasers	0.55
	No rulers	0.60
	No pens or ball point pens	0.62

Factor	Variables	Loadings
TBEHAVE		
	Arrive late	0.57
	Absenteeism	0.68
	Skip class	0.67
	Bully pupils	0.58
	Sexually harass pupils	0.54
	Language	0.64
	Drug abuse	0.45
	Alcohol abuse	0.65
	Health problem	0.46
RESOURCE		
	Library	0.40
	Hall	0.49
	First aid	0.44
	Electricity	0.72
	Telephone	0.73
	Fax	0.61
	Typewriter	0.57
	Duplicator	0.76
	Tape recorder	0.54
	TV	0.75
	VCR	0.72
	Photocopier	0.62
	Computer	0.69
	Cafeteria	0.36
COMUNITY		
	Build facility	0.68
	Maintain facility	0.86
	Furniture equipment	0.87
	Textbooks	0.75
	Stationery	0.83
	Other materials	0.83
	Exam fees	0.63
	Staff salary	0.53
	Extra curricular	0.50

Note:
Factor - Principal component factor