J. R. Statist. Soc. A (2007)
170, *Part* 4, *pp.* 941–954

Modelling the effect of pupil mobility on school differences in educational achievement

Harvey Goldstein, Simon Burgess and Brendon McConnell
University of Bristol, UK

[Received September 2006. Final revision February 2007]

Summary. The recently introduced national pupil database in England allows the tracking of every child through the compulsory phases of the state education system. The data from key stage 2 for three local education authorities are studied, following cohorts of pupils through their schooling. The mobility of pupils among schools is studied in detail by using multiple-membership multilevel models that include prior achievement and other predictors and the results are compared with traditional 'value-added' approaches that ignore pupil mobility. The analysis also includes a cross-classification of junior and infant schools attended. The results suggest that some existing conclusions about schooling effects may need to be revised.

Keywords: Cross-classified model; Educational attainment; Mobility; Multilevel model; Multiple-membership model; National pupil database; Pupil level annual school census; Random effects; School effectiveness; Value added; Variance components

1. Introduction

Since the early 1980s educational researchers have developed models for judging the comparative performance of schools and other institutions by using what have come to be known as 'value-added' techniques (see Goldstein *et al.* (1993) and Raudenbush and Bryk (1989) for early discussions). Typical applications have compared the performance of pupils in public examinations or on the basis of routine test scores. In essence these models attempt to adjust simple comparisons of school mean values by using measures of pupil prior achievement and other variables to take account of selection and other procedures that are associated with pupils' achievements but not related to any effect that the schools themselves may have on achievement. Thus, a simple two-level, variance components, model based on data from a random sample of schools can be written as follows where subscript i refers to the pupil, and j to the school:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_j + e_{ij}, \qquad u_j \sim N(0, \sigma_u^2), \quad e_{ij} \sim N(0, \sigma_e^2); \tag{1}$$

here y_{ij} and x_{ij} respectively are the response variable and prior achievement, and u_j is an underlying school effect or residual (which is associated with school organization, teaching, etc.). As is usual in models of this kind, we assume that e_{ij} and u_j are uncorrelated and also uncorrelated with any explanatory variables—i.e. we assume that any possible dependences that may result from, for example, school selection mechanisms are accounted for. Posterior estimates $\hat{u}_j$ with associated confidence intervals are typically used to rank schools in so-called 'league

Address for correspondence: Harvey Goldstein, Centre for Multilevel Modelling, University of Bristol, 2 Priory Road, Bristol, BS8 1TX, UK.
E-mail: h.goldstein@bristol.ac.uk

 0964–1998/07/170941

tables' or used as 'screening devices' in school improvement programmes (Goldstein *et al.*, 2000a).

Model (1) can be elaborated by introducing further covariates such as socio-economic background or peer group characteristics, to make additional adjustments, satisfy the distributional assumptions or investigate interactions. In addition, it is typically found that models such as model (1) require random coefficients where, for example, the coefficient of prior achievement varies randomly across schools. In this case, using a more general notation, we have

$$y_{ij} = \beta_{0ij} + \beta_{1j} x_{ij},$$

$$\beta_{0ij} = \beta_0 + u_{0j} + e_{ij},$$

$$\beta_{1j} = \beta_1 + u_{1j},$$

$$e_{ij} \sim N(0, \sigma_e^2), \qquad \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N(0, \Omega), \qquad \Omega = \begin{pmatrix} \sigma_{u0}^2 & \\ \sigma_{u01} & \sigma_{u1}^2 \end{pmatrix}.$$

These 'multilevel' models have also been extended to include further levels of hierarchy, such as education board or authority and random factors which are not contained within a simple hierarchy such as area of pupil residence or school attended during a previous phase of education. Such designs are known as 'cross-classifications'. In the present paper we incorporate a cross-classification for a previous phase of education, namely infant school attended when studying differences between junior schools. We also incorporate a 'multiple-membership' structure to take account of the fact that many pupils change schools so that in terms of the 'effect' of school more than one school will contribute. We also study the effects of further covariates measured on pupils and schools, including 'compositional' effects that are aggregates of student level variables such as whether or not a pupil is eligible for free school meals. The full model including a cross-classification of infant and junior schools with multiple membership of junior schools can be written as

$$\left.\begin{array}{l} y_i = \beta_0 + \sum_h X_{hi}\beta_h + u^{(1)}_{\text{infant}(i)} + \sum_{j \in \text{junior}(i)} w^{(2)}_{i,j} u^{(2)}_j + e_i, \\ \sum_{j \in \text{junior}(i)} w^{(2)}_{i,j} = 1, \qquad \text{infant school}(i) \in (1, \ldots, J_1), \qquad \text{junior school}(i) \in (1, \ldots, J_2), \\ u^{(1)}_{\text{infant}(i)} \sim N(0, \sigma^2_{u(1)}), \qquad u^{(2)}_{\text{junior}(i)} \sim N(0, \sigma^2_{u(2)}), \qquad e_i \sim N(0, \sigma^2_e), \qquad i = 1, \ldots, N. \end{array}\right\} \tag{2}$$

In model (2) we have used a general notation that is sufficiently flexible to describe our models. The superscripts in model (2) refer to the set of units for each, independent, cross-classification: (1) refers to infant schools and (2) to junior schools. The weights $w^{(2)}_{i,j}$ sum to 1 for each pupil so that the term

$$\sum_{j \in \text{junior}(i)} w^{(2)}_{i,j} u^{(2)}_j$$

represents the weighted contribution from the junior schools attended to the score for pupil i. Each infant school has a random effect and each junior school has a random effect with independent normal distributions. Goldstein (2003) gives further details and describes estimation procedures.

A major aim of our analyses is to see whether taking account of the cross-classified and multiple-membership nature of pupil achievement data substantially alters inferences in current analyses of school effects, the coefficients associated with covariates and the relative values of the school effect estimates themselves, $\hat{u}_j$. Although inferences of a fully causal nature are important in these kinds of analyses, our principal concern is to judge whether these altered inferences have important implications for school rankings, and the relative amounts of variation that are ascribed to different types of school. Nevertheless, in terms of the causal classification that was suggested by Raudenbush and Willms (1995), model (1), with its associated assumptions, is essentially the model for their 'type A' effects, i.e. the total set of influences on pupils' progress after adjusting for appropriate covariates such as prior achievement.

The analyses that we describe have only recently become possible with the release of the pupil level annual school census (PLASC) data, which we describe below. The importance of this work is related to the fact that the literature on the effects of schooling (see Goldstein (2001) for a review) has largely ignored the existence of cross-classifications and pupil mobility. In addition, educational policy in England and in certain other education systems uses school effect estimates as part of a public accountability system that also ignores these issues.

The next section describes the data set and variables that are used and is followed by the results of fitting a series of cross-classified and multiple-membership models.

2. The pupil level annual school census data set

In the following analysis, we use the national pupil database (NPD) that includes data from key stage (KS) tests and the PLASC data set from the Department for Education and Skills. The database covers all pupils in state primary and secondary schools in England and can be linked to each pupil's test score history. In addition, it contains some personal and school characteristics: ethnicity, gender, within-year age, mother tongue, an indicator of family poverty (eligibility for free school meals, which is dependent on receipt of some welfare benefits) and an indicator of special educational needs. At the time of analysis, 3 years of data had been released: 2002, 2003 and 2004. For more details see `http://www.bris.ac.uk/Depts/CMPO/PLUG/whatisplug.htm`.

From these data sets, we take a cohort of pupils who took their KS1 examinations in 2000 at the end of their second year and their KS2 examinations in 2004 at the end of year 6. In the English school system children enter the reception year during the academic year when they reach the age of 4 years. The following 2 years (years 1 and 2) are spent in the infant department or section, either in a separate school or as part of a primary school. The following 4 years (years 3–6) are spent in the junior department, either in a separate school or as part of a primary school.

We know which school they were in at five time points—the times of the two KS tests, as well as the three census dates in the PLASC—January 2002, 2003 and 2004. We consider the mobility of pupils in the last 3 years of primary school, in the junior department, and cross-classify them by their infant department or school.

2.1. Sample selection

The NPD contains only data on state school pupils. Consequently, these results will underestimate mobility as we shall lose pupils who migrate to the private sector part way through primary school. Pupils are also dropped if they had missing school identification data at any of the five stages. This meant losing just under 5% of the data (28 687 out of an initial 586 622 pupils). Again this is likely to underestimate pupil mobility, as some of the missing data may be due to pupils moving between schools. Finally, pupils in schools with five or fewer pupils within the cohort were dropped. This dropped 0.4% of pupils (2264/557935), resulting in 555671 for analysis.

We chose to restrict the analysis to three particular local education authorities (LEAs). This means that pupils were included only if they took their KS2 examinations in the LEA of interest, wherever their previous school was, and not if they simply ever attended a school within this LEA. Initially, Hampshire, Northamptonshire and Staffordshire were chosen as LEAs, being of a similar size. The last two have rather more pupil mobility than the first where only 9% moved and accounting for this movement has a negligible effect on estimates. For the present paper, therefore, we use data only from Northamptonshire and Staffordshire.

2.2. Pupil movement

Looking at the national data, 15% of pupils moved between the PLASC census date of January 2002 and the KS2 test date in 2004, and 43% between KS1 in 2000 and the KS2 test date. Since PLASC data collection started only in 2002 we can use detailed information about movements for only years 4, 5 and 6. Table 1 shows the patterns of pupil movement in the final 3 years of primary school where we have taken the January 2002 status as equivalent to that at the start of school year 4. The patterns allow for up to four schools to be attended, with the lengths of time in each indicated. The final time of 0.6 refers to the period between the final PLASC census date in year 6 (PLASC 2004) and the time of the KS2 test. From this the variable pattern with eight categories is constructed to give an overview of time spent in each school, with the time spent in the first school listed first.

Mobility differs across the three LEAs that were considered—for the period between PLASC 2002 and KS2, the proportion of those in the same school were Northamptonshire 61%, Staffordshire 75% and Hampshire 91%.

The pattern of mobility for each of the LEAs, corresponding to the patterns in Table 1, is shown in Table 2, where just the percentages are given.

3. Variables used in the analysis

The data were selected from the NPD, recoded by using Stata version 8 (`http://www.stata.com`) and then input to MLwiN (Rasbash *et al.*, 2004) via EXCEL for the model fitting. The student level variables that were used are set out in Table 3.

Table 1. Time spent in the 3-year period covered by years 4 (PLASC 2002), 5 and 6†

Pattern	*Frequency*	*%*
1, 1, 0.4, 0.6	208	0.04
1, 1, 1	5017	0.90
1, 1.4, 0.6	541	0.10
1, 2	53274	9.59
2, 0.4, 0.6	493	0.09
2, 1	24131	4.34
2.4, 0.6	2144	0.39
3	469863	84.56
Total	555671	100.00

†National data.

Table 2. Mobility percentages for three LEAs†

Pattern	*Hampshire (%)*	*Staffordshire (%)*	*Northamptonshire (%)*
1, 1, 0.4, 0.6	0.02	0.01	0.05
1, 1, 1	0.65	0.70	8.22
1, 1.4, 0.6	0.06	0.05	0.16
1, 2	4.59	20.89	27.33
2, 0.4, 0.6	0.09	0.05	0.03
2, 1	3.27	2.69	2.95
2.4, 0.6	0.31	0.25	0.35
3	91.00	75.36	60.92
Total	100.00	100.00	100.00
Sample size	13698	9581	7640

†The definitions are as for Table 1.

Table 3. Student level variables that were used in the analysis†

Variable	*Description*
Gender	Male or female
First language	Whether or not English is the first language of the child, as recorded in PLASC 2004
Ethnicity	Provided by child, school or parent; recoded as ‘white’, ‘Asian’, ‘black’, ‘Chinese’, ‘mixed’ or ‘other’
Number of moves	Number of schools the child has moved during the 3-year period years 4–6
Free school meal eligibility	Eligibility for free school meals recorded at PLASC 2002 and 2004 and recoded; free school meals in 2002 and whether same status at 2004, moved into free school meals eligibility by 2004, moved out of free school meals eligibility by 2004
KS1 test score	KS1 mathematics test score at end of year 2
Age in months	Age of the child in months from the end of the school year so that the youngest child born in August is coded 1 and the oldest born in September is coded 12; there is a very small number (less than 0.25%) not in the modal year group and we have not treated these separately since there are too few.
Special educational needs	Two variables: whether the child was statemented in 2002 and whether statemented in 2004; the term ‘statemented’ refers to a formal statement for children with learning difficulties about their special educational needs; it includes 2–3% of the population
Infant or junior school category KS1	Type of infant school the pupil is in at KS1 test: single primary school; separate infant and junior schools on the same site; separate infant and junior schools on different sites

†Pupil test scores have been normalized (see the text). For each student the date that they entered or left a school is recorded.

The school level variables that were used in the analyses are described in Table 4.

Other aggregated variables were used in exploratory analyses but not found to contribute to prediction or explanation and are not included here. The original scores on the tests are derived from a series of discrete ‘levels’ ascribed to the results (see `http://www.dfes.gov.uk/performancetables` for a description). So that we can more closely approximate the distributional assumptions of our models, where they are used at the pupil level we have

Table 4. School level variables that were used in the analysis, aggregated from pupil level data†

Variable	*Description*
KS1 mean mathematics score for KS2 cohort	Mean KS1 mathematics score of the children in the school where the index child takes their KS2 test
KS1 mean mathematics score for KS1 cohort	Mean KS1 mathematics score of the children in the school where the index child takes their KS1 test
KS1 standard deviation of mathematics score for KS1 cohort	Standard deviation KS1 mathematics score of the children in the school where the index child takes their KS1 test
% eligible for free school meals at KS2	% of pupils in the index child's school at KS2 eligible for free school meals at KS2
% eligible for free school meals at KS1	% of pupils in the index child's school at KS1 eligible for free school meals at KS1
% white at KS2	% of pupils in the index child's school at KS2 who are white ethnic

†Mean KS1 mathematics scores are not normalized.

monotonically transformed the total KS1 and KS2 distributions to normality by assigning equivalent points on the standard normal distribution scale. The analyses have been carried out for mathematics test scores only at KS2 (response) and KS1.

We also note that for the cohort mean variables, strictly speaking, we should use a weighted average over all the 'peer groups' that the child has been with for the KS1–KS2 period, since this would be expected better to reflect the actual peer group influences. When we use the mean (or standard deviation) KS1 scores for the pupils in the 'target' child's school at the time of the KS1 test this is in fact an approximation. Ideally we would like to measure the mean KS1 score of the pupils in the target child's school during year 2, weighted according to the time that each pupil has been in that school. In addition, if the target child has moved schools we would have to modify the computations accordingly to include the pupils in every school attended. The data that are available for the year 2000 do not allow this; we have only the school attended and test score at the time of the KS1 test. As future data become available, however, such computations will become possible.

Likewise, we might expect that continuing peer group effects could be important for the years between the KS1 and KS2 tests. Thus, for years 4, 5 and 6 we could also calculate the mean KS1 test scores for all the pupils in the same school as the target child, using suitable weights. In addition, we require a further set of multiple-membership weights related to the length of time that the target child spends in each school. For each target child, combining these two weightings, we would derive a composite weight attached to each pupil who spends any time in the same school with the target child, and these would be used to derive the mean KS1 score. Although these latter computations are technically possible with the current PLASC data, they are computationally demanding and there seems little point in so doing in the absence of the year 2 (PLASC year 2000) data.

4. Data analysis

In MLwiN the data are stored by pupil record. For the multiple-membership analyses MLwiN requires each pupil to have up to q columns reserved containing the school identification codes for each school attended together with a 'weight' that is used in the analysis. For each pupil, the ordering of the schools attended is unimportant, so long as the correct weights are attached. MLwiN requires that the first data set column containing the first-school identifier for each pupil comprises a superset of all the school identifiers. The different weighting systems are described below. A description of how to set up such an analysis is given in Browne (2004), chapter 15. The first set of variance component analyses are fitted by using traditional maximum likelihood estimation. The results are almost identical to those which were obtained by using Markov chain Monte Carlo (MCMC) sampling to which we switch for the cross-classified and multiple-membership models.

The analysis is first described in detail for Staffordshire and results are presented for Northamptonshire, noting differences and similarities. We fit models of increasing complexity.

For Staffordshire we have 9226 pupils with both KS1 and KS2 scores. For the traditional two-level value-added analysis where a pupil is assigned just to the school in which they take their KS2 test we have 241 schools. Adding in all the schools the pupils attended both within and outside Staffordshire we have 591 junior schools or departments and 769 infant schools or departments. For Northamptonshire the corresponding numbers are 7329 pupils, 183 schools in the traditional value-added analyses and 505 junior and 683 infant schools or departments.

4.1. Variance component analyses using key stage 2 school identifier

The first analyses that are shown in Table 5, for Staffordshire, fit KS2 mathematics test score as response with, successively, an intercept, pupil age and KS1 test score. The level 2 identifier is the KS2 school.

The variance partition coefficient VPC (Goldstein, 2003), i.e. the proportion of variance at the school level, is 11% for the simple model where just an intercept is fitted (model A).

The pupil's age at this stage of schooling is now added to the basic model since it is known that age is related to attainment at this stage of schooling and our analysis confirms this (model B). We see a positive relationship, as expected, whereby there is a difference of nearly a third of a standard deviation (0.029×11) in mathematics test score between the oldest and youngest pupils in the year. Model C is the most common or traditional value-added analysis where we additionally condition on the KS1 mathematics score by using a linear term which is sufficient to describe the relationship.

Table 5. Normalized mathematics KS2 score response for Staffordshire, with pupils assigned to KS2 test score school by using models of increasing complexity†

Variable	*Results for the following models:*		
	A	*B*	*C*
Fixed effects			
Intercept	−0.011	−0.195	−0.093
Age in months		0.029 (0.003)	−0.014 (0.002)
KS1 mathematics score			0.754 (0.006)
Random parameters			
Between-junior-school variance	0.094 (0.011)	0.095 (0.011)	0.053 (0.006)
Between-pupil variance	0.795 (0.012)	0.784 (0.012)	0.301 (0.004)
VPC	0.11	0.11	0.15
Deviance (−2 log-likelihood)	25107.1	24989.0	15987.9

†Estimates with standard errors in parentheses.

As expected, there is a substantial decrease in the variances at both levels moving to model C and VPC increases from 11% to 15%. The relationship with age in the value-added model is now negative so that given their KS1 performance the younger children do better, indicating that they tend to 'catch up' over this period. See also Goldstein and Fogelman (1974) for a similar finding.

The following analyses all include KS1 score so the remaining effects can be interpreted in terms of affecting the change between KS1 and KS2 scores, which can be interpreted as a measure of progress. The next analysis summarizes several explorations fitting different combinations of variables and retains both those statistically significant and those which have some substantive interest.

We see that there is no significant effect for English as a first language and the only ethnic group which shows a difference from the white group (which is used as the base category) is Chinese. Boys show greater progress than girls, those who move more often tend to show less progress, being eligible for free school meals in year 4 (2002) is associated with less progress as is moving into that category between year 4 and year 6 and statemented pupils show considerably less progress than non-statemented pupils. A higher percentage of free school meal pupils in the pupil's KS1 school is correlated with lower progress made and additionally the higher the

Table 6. Mathematics KS2 score response for Staffordshire, with pupils assigned to KS2 test score school: traditional value-added model†

Variable	*Estimate*	*Standard error*
Fixed effects		
Intercept	2.255	
Age in months	−0.014	0.002
KS1 mathematics score centred	0.756	0.007
Not English first language − English first language	0.048	0.092
Mixed ethnic − white	0.050	0.047
Asian − white	0.038	0.088
Black − white	−0.068	0.104
Chinese − white	0.524	0.160
Other ethnic − white	0.078	0.131
Number of school moves	−0.052	0.020
Male − female	0.089	0.011
Moved into free school meals year 4 to year 6 − no free school meals	−0.101	0.036
Moved out of free school meals year 4 to year 6 − no free school meals	0.022	0.035
Free school meals in year 4 and year 6 − no free school meals	−0.083	0.022
Statemented in year 6 − non statemented	−0.267	0.037
Proportion in pupil's KS1 class with free school meals	−0.795	0.097
Proportion in pupil's KS2 class with free school meals	−0.620	0.158
KS1 mean KS1 score in pupil's KS1 school	−0.118	0.009
KS1 standard deviation of KS1 scores in pupil's KS1 school	−0.051	0.014
Random parameters		
Between-junior-school variance	0.039	0.004
Between-pupil variance	0.292	0.004
VPC	0.12	
Deviance (−2 log-likelihood)	15359.1	

†All variables labelled as differences, e.g. 'black − white', are contrasts between the two categories shown. The analysis uses corresponding dummy variables and the base categories are English first language, white, female, no free school meals and non-statemented.

percentage in the pupil's KS2 school the less progress made. The higher the mean KS1 mathematics score of the pupils in the pupil's KS1 school the less progress is made and the greater the variability there the less progress is made. This result has been found elsewhere and suggests that there is a complex interaction between the pupil's score and that of their peers, but we have not investigated this in detail (see Bryk and Raudenbush (2002) for a further discussion).

If we allow the coefficient of the KS1 mathematics score to vary across schools, a random-coefficient model, the fixed coefficient estimates are little changed, but we do obtain a significant between-school variance in the coefficient of KS1 mathematics score.

Since the main purpose of the present paper is to explore the effect of introducing multiple-membership and cross-classified models we shall not follow up this evidence here of 'differential effectiveness', but it will be explored in a subsequent publication, especially in terms of the estimation of 'school effects'. The following analyses therefore will use only a variance components model. We have also explored the type of junior school or department attended but this variable is not significant.

In the following analyses we use MCMC estimation with a burn-in of 500 and a chain of 5000. Diffuse priors are used, as described in Browne (2004). We switch our estimation method to MCMC sampling because for the multiple-membership and cross-classified data models it is computationally very much more efficient. A discussion of this is given by Browne *et al.* (2001).

Table 7. Mathematics KS2 score response for Staffordshire: MCMC estimates with multiple-membership structure

Variable	*Estimate*	*Standard error*
Fixed effects		
Intercept	2.268	
Age in months	−0.014	0.002
KS1 mathematics score centred	0.757	0.007
Not English first language − English first language	0.059	0.093
Mixed ethnic − white	0.051	0.047
Asian − white	0.034	0.088
Black − white	−0.065	0.105
Chinese − white	0.504	0.163
Other ethnic − white	0.066	0.133
Number of school moves	−0.055	0.021
Male − female	0.089	0.011
Moved into free school meals year 4 to year 6 − no free school meals	−0.099	0.036
Moved out of free school meals year 4 to year 6 − no free school meals	0.025	0.035
Free school meals in year 4 and year 6 − no free school meals	−0.085	0.023
Statemented in year 6 − non statemented	−0.263	0.038
Proportion in pupil's KS1 class with free school meals	−0.798	0.100
Proportion in pupil's KS2 class with free school meals	−0.683	0.154
KS1 mean KS1 score in pupil's KS1 school	−0.118	0.009
KS1 standard deviation of KS1 scores in pupil's KS1 school	−0.056	0.015
Random parameters		
Between-junior-school variance	0.047	0.006
Between-pupil variance	0.291	0.004
VPC	0.14	
DIC	15144.2	

Using MCMC estimation for the model in Table 6 we obtain estimates that are very similar to the maximum likelihood estimates. The deviance information criterion DIC has a value of 15182.0 for this model. This is a measure of model complexity and we shall use it to compare models (see Spiegelhalter *et al.* (2002)).

4.2. Multiple-membership and cross-classified models

We now incorporate the movements of pupils between schools by extending the model in Table 6 to a multiple-membership model as given by model (2) without the infant school effects. The weights are proportional to the time that a pupil spends in each school. Because the PLASC data cover the whole of England we have many pupils in Staffordshire schools at the time of the KS2 test who were in schools outside Staffordshire at some time between the start of year 4 and the KS2 tests, and these schools can be identified. In fact there are a total of 591 distinct schools in the data set of which only 241 are within Staffordshire. For most of these extra schools only one pupil is represented in the data set.

When the multiple-membership model is fitted we obtain the results in Table 7.

The DIC value is 15144.2 compared with a value of 15182.0 before—a substantial reduction. The main effect has been to increase the school level variance by 18%, leaving the other

Table 8. Mathematics KS2 score response for Staffordshire: MCMC estimates with multiple-membership structure for junior crossed with infant

Variable	*Estimate*	*Standard error*
Fixed effects		
Intercept	2.173	
Age in months	−0.014	0.002
KS1 mathematics score centred	0.757	0.007
Not English first language − English first language	0.070	0.091
Mixed ethnic − white	0.042	0.046
Asian − white	0.037	0.087
Black − white	−0.062	0.104
Chinese − white	−0.500	0.158
Other ethnic − white	0.071	0.135
Number of school moves	−0.056	0.020
Male − female	0.090	0.011
Moved into free school meals year 4 to year 6 − no free school meals	−0.094	0.036
Moved out of free school meals year 4 to year 6 − no free school meals	0.021	0.036
Free school meals in year 4 and year 6 − no free school meals	−0.084	0.022
Statemented in year 6 − non statemented	−0.264	0.037
Proportion in pupil's KS1 class with free school meals	−0.812	0.113
Proportion in pupil's KS2 class with free school meals	−0.528	0.149
KS1 mean KS1 score in pupil's KS1 school	−0.113	0.010
KS1 standard deviation of KS1 scores in pupil's KS1 school	−0.053	0.010
Random parameters		
Between-junior-school variance	0.012	0.004
Between-infant-school variance	0.038	0.005
Between-pupil variance	0.288	0.004
VPC: junior	0.04	
VPC: infant	0.11	
DIC	15107.0	

effects relatively unaltered. Some discussion of this will be given in our later concluding remarks.

An issue with the multiple-membership analysis is the choice of weights. We have investigated several choices, including giving more or less weight to periods that are spent in schools further away in time from the KS2 test. If we weight the time that is spent in each school by the time difference between the mid-point of the period in the school and the time of KS2 this gives more weight to the earlier schools and we obtain a DIC of 15158 which is greater than that above. Simply giving equal weight to every school that the pupil attends gives a DIC of 15146 which is close to the choice of weights proportional to time spent and we shall use this weighting system in subsequent analyses.

We now look at the effect of including the infant school or department attended. Since every pupil attends both an infant and a junior department or school there is a cross-classification here and the next model, with results in Table 8, incorporates both this cross-classification and the multiple membership.

The DIC value of 15107.0 is a considerable reduction as would be expected. Most noticeably, the infant variance is three times as large as the junior variance, even though we have not been able to fit a multiple-membership model for infants. This finding that earlier school membership contributes more variation than later school membership is echoed by other analyses (Goldstein and Sammons, 1997) at different stages of schooling, but, as we shall see, this is not repeated in Northamptonshire. We can also look at the estimates of school effects, the level 2 residuals, and Fig. 1 compares these estimates for the simple 'value-added' model using KS2 school as identifier with the full multiple-membership cross-classified model. The estimates of the fixed effects and the level 1 variance are also very little changed across analyses.

The residuals are very highly correlated (0.98) and they also have similar standard errors. This suggests that at the junior stage of education movement between schools is independent of school effects and that classifying schools by ignoring both the multiple-membership and the cross-classified structures will not result in any serious change in rankings.

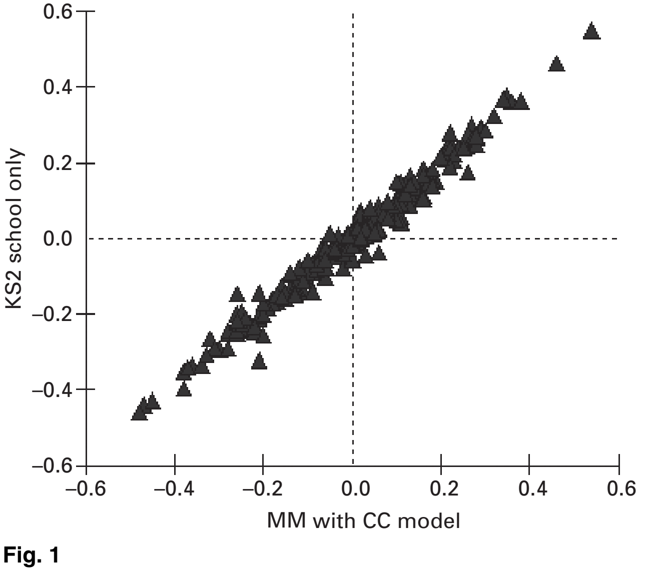

Fig. 1

Table 9. Mathematics KS2 score response for Northamptonshire: cross-classified and multiple-membership model

Variable	*Estimate*	*Standard error*
Fixed effects		
Intercept	−0.010	
Age in months	−0.009	0.002
KS1 mathematics score centred	0.740	0.007
Not English first language − English first language	0.227	0.064
Asian − white	−0.130	0.071
Black − white	0.042	0.051
Other ethnic − white	−0.041	0.042
Number of school moves	−0.042	0.019
Male − female	0.078	0.013
Moved into free school meals year 4 to year 6 − no free school meals	−0.152	0.040
Moved out of free school meals year 4 to year 6 − no free school meals	0.082	0.040
Free school meals in year 4 and year 6 − no free school meals	−0.139	0.028
Statemented in year 6 − non statemented	−0.321	0.069
Proportion in pupil's KS1 class with free school meals	−0.191	0.106
Proportion in pupil's KS2 class with free school meals	−0.403	0.169
Transition from infant school same site as junior − infant department or section of primary school	−0.030	0.041
Transition from infant school different site from junior − infant or section of primary school	−0.060	0.040
Random parameters		
Between-junior-school variance	0.035	0.006
Between-infant-school variance	0.023	0.005
Between-pupil variance	0.280	0.005
VPC: junior	0.10	
VPC: infant	0.07	
DIC	11750.0	

We now look at the corresponding analyses for Northamptonshire. We present only the final set of multiple-membership and cross-classified analyses, omitting non-significant (at the 5% level) explanatory variables. For ethnic group we can only distinguish Asian, black, white and other. We include the infant–junior transition type since this has a significant effect.

Fitting both the multiple-membership and the cross-classification structure we obtain the results in Table 9.

Unlike in the case of Staffordshire, the infant school variance is now smaller than the junior school variance and this is still so if we do not include type of infant school. As before, however, the correlation between junior school effects for the simple and the full models is very high at 0.97. We should note that there are 21 middle schools, where pupils transfer at year 5, out of 246 schools for pupils in years 4–6. These schools are not present at the time of the KS2 test and so function as if they are schools from outside the LEA.

If we fit a cross-classified model but without multiple membership, and using KS2 school identification, the between-junior and between-infant variances are respectively 0.024 and 0.028 with similar standard errors as in Table 9. Thus, unlike Staffordshire, the variances that are associated with the two stages of schooling are similar.

5. Conclusions and further work

In 'school effectiveness' studies of the contribution of schools to educational outcomes, pupils' test scores are typically assigned to a single school, usually the school in which the outcome measure, a test or examination, is taken. In fact, pupils spend many years in school, their learning is cumulative and there is movement between schools both within a particular stage (e.g. the junior school period) and across stages (e.g. from the infant to junior stage). Hence, assignment of a pupil's test or examination score to a single school is an approximation that may distort inferences about the effects of schooling. In this paper we have examined the effect of allowing for such movement on parameter estimates. We use recently available data to do this (the PLASC and NPD) that allow us to track pupils across schools. The two issues that we focus on are whether this allowance for mobility influences rankings of schools ('league tables') and whether it substantially changes the estimate of the between-school variation.

We have two main results. First, we show that the traditional value-added model, that ignores mobility, underestimates the importance of the school as measured by its contribution to the overall variance. We have also shown that the relationships between variances at different stages of schooling are changed. This suggests that many of the conclusions about school effects in the educational literature may need revision and that future studies should be designed so that pupil mobility is properly accounted for. The reason for the upward revision of the between-school variance is because when mobility is ignored the estimate that is obtained has the variance associated with an average over several schools (those actually attended) which is smaller than the true between-school variance. In the multiple-membership models one issue is the choice of a weighting system. Our limited exploration suggests that a simple system that defines weights proportional to the time spent in each school is near optimum.

The present paper adds to a recent literature re-evaluating the importance of school and teacher variation in influencing test score outcomes. Rockoff (2004), Kane and Staiger (2005), Rivkin *et al.* (2005) and Aaronson *et al.* (2007) all found the learning environment—teachers and schools—to be considerably more important than previous results suggested, once appropriate allowance has been made for the potentially non-random assignment of pupils to schools and teachers to classes. It is notable that those advances, as with this paper, rely on much better data to be able to make headway on this problem.

Secondly our analyses suggest that, for the purpose of ranking schools on the basis of their posterior value-added estimates, no serious errors will be made by ignoring pupil mobility. In addition, the use of cross-classified models does provide further insight into the effects of prior stages of schooling, although this also does not appear to alter the rankings of the estimated school effects. We should, however, make the *caveat* that our conclusions may be modified when random-coefficient, 'differential effectiveness' models are used that allow the coefficients of prior achievements to vary across schools. We further note that the existence of such random coefficients, when ignored in a variance components model, may induce an association between the school level residuals and the prior achievement covariate.

In future work we intend to introduce area of residence as another classification and additionally to take account of movements among areas. We shall also extend the work to include the whole of the English educational system. Thus when modelling multiple membership we shall not be forced to include a very large number of schools with just one or two pupils. This will, however, impose a considerable computational burden and we shall be exploring ways of coping with this, possibly through a carefully constructed sampling scheme.

Our procedures also apply to areas other than education, e.g. in repeated measures designs that are treated as two-level structures with measurement occasions nested within individuals.

Thus, in a study where there are measurer effects and measurers change over time, a multiple-membership model is required. Another example is in panel studies of households where there is movement of individuals between households over time and where the use of multiple-membership models is necessary for certain types of question (Goldstein *et al.*, 2000b). In all these cases one of the major problems is the availability of data that allow mobility to be tracked and this suggests that efforts should be made at the study design stage to ensure that these are collected.

Acknowledgements

We are most grateful to Anna Barker and Catherine Blackham of the Department for Education and Skills for support and help with the data sets and to referees and the journal Joint Editor for helpful comments. This work was partly supported by the Economic and Social Research Council (grant RES-500-25-5002) and the Department for Education and Skills.

References

Aaronson, D., Barrow, L. and Sander, W. (2007) Teachers and student achievement in the Chicago public high schools. *J. Lab. Econ.*, **25**, 95–136.
Browne, W. J. (2004) *MCMC Estimation in MLwiN, version 2.0*. London: Institute of Education.
Browne, W., Goldstein, H. and Rasbash, J. (2001) Multiple membership multiple classification (MMMC) models. *Statist. Modllng*, **1**, 103–124.
Bryk, A. S. and Raudenbush, S. W. (2002) *Hierarchical Linear Models*. Newbury Park: Sage.
Goldstein, H. (2001) Using pupil performance data for judging schools and teachers: scope and limitations. *Br. Educ. Res. J.*, **27**, 433–442.
Goldstein, H. (2003) *Multilevel Statistical Models*, 3rd edn. London: Arnold.
Goldstein, H. and Fogelman, K. (1974) Age standardisation & seasonal effects in mental testing. *Br. J. Math. Statist. Psychol.*, **44**, 109–115.
Goldstein, H., Huiqi, P., Rath, T. and Hill, N. (2000a) *The Use of Value Added Information in Judging School Performance*. London: Institute of Education.
Goldstein, H., Rasbash, J., Browne, W., Woodhouse, G. and Poulain, M. (2000b) Multilevel models in the study of dynamic household structures. *Eur. J. Popln*, **16**, 373–387.
Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., Pan, H., Nuttall, D. and Thomas, S. (1993) A multilevel analysis of school examination results. *Oxf. Rev. Educ.*, **19**, 425–433.
Goldstein, H. and Sammons, P. (1997) The influence of secondary and junior schools on sixteen year examination performance: a cross-classified multilevel analysis. *School Effect. School Improvmnt*, **8**, 219–230.
Kane, T. J. and Staiger, D. O. (2005) Using imperfect information to identify effective teachers. *Working Paper*. Dartmouth College, Hanover.
Rasbash, J., Steele, F. and Browne, W. (2004) *A User's Guide to MLwiN Version 2.0*. London: Institute of Education.
Raudenbush, S. W. and Bryk, A. S. (1989) Quantitative models for estimating teacher and school effectiveness. In *Multilevel Analysis of Educational Data* (ed. R. D. Bock). New York: Academic Press.
Raudenbush, S. W. and Willms, J. D. (1995) The estimation of school effects. *J. Educ. Behav. Statist.*, **20**, 307–336.
Rivkin, S. G., Hanushek, E. A. and Kain, J. F. (2005) Teachers, schools and academic achievement. *Econometrica*, **73**, 417–458.
Rockoff, J. E. (2004) The impact of individual teachers on student achievement: evidence from panel data. *Am. Econ. Rev.*, **94**, 247–252.
Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002) Bayesian measures of model complexity and fit (with discussion). *J. R. Statist. Soc.* B, **64**, 583–639.