To appear in Oxford Review of Education, 1999

The use of assessment data for school improvement purposes

By

Min Yang*
Harvey Goldstein*
Terry Rath**
Nigel Hill**

**Institute of Education, University of London*
***Hampshire Local education Authority*

Abstract

Hampshire LEA has carried out two longitudinal studies: from reception intake to the end of year 2, and from the end of year 2 to the end of year 6. A total, respectively, of 161 and 114 schools and about 6400 and 4700 pupil records have been analysed. Test scores at baseline (entry to reception), key Stage 1 and key stage 2 have been used, together with pupil and school level variables. Multilevel models have been fitted and show that schools differ along several dimensions, both by curriculum subject and by prior attainment of pupil.

The research sought ways of communicating the results to head teachers in ways which were meaningful without destroying the underlying complexity of the relationships uncovered. The paper describes how this can be done, in ways which can assist the process of school improvement.

Keywords

Educational effectiveness, school improvement, value added, multilevel models, statistical adjustment, assessment, baseline assessment.

Acknowledgements

We would like to acknowledge the head teachers who took part in this research; for their enthusiasm and their contributions to the use of the results for school improvement.

Introduction

The systematic publication of 'performance tables' for public examination results, begun in 1992, is now an established feature of the educational system in England and Wales. In 1996 the Government has published such tables for key stage 2 test scores, at both local and national level and these will continue (D*f*EE, 1997). These tables contain average, unadjusted, test scores for each school. At the same time the Government has indicated that it wishes to see the publication of so called 'value added' or 'intake adjusted' scores and a project under the auspices of the Qualifications and Curriculum Authority (QCA, 1997) has studied the implementation of such a scheme nationally.

The principal argument against unadjusted 'league tables' is that the performance of a school is determined largely by the pre-existing achievements of the students when they enter it. Since schools differ markedly in this respect, for example some schools are highly selective - either deliberately or indirectly for socio-demographic reasons, it is impossible to judge the quality of the education *within a school* solely in terms of such outputs. Nevertheless, there are also problems which apply to 'value added' tables, and it was in order to explore the limitations as well as the potential of these that the project to be described in this paper was carried out. For a technical discussion of the issues surrounding the analysis and interpretation of institutional performance data see Goldstein and Spiegelhalter (1996).

Data sources

Baseline 1993 to Key stage 1 1996 (BKS1)

The dataset contains records of 6907 pupils from 169 schools. The analysis has been based on the records for about 6400 pupils in 161 schools. Data are missing from schools on the student level variables SEN (*statement for special educational need*, 1 school lost), SENAUDIT (*stages for special education need*, 3 schools lost) and ABSENCE (*Number of half days absent*, 4 schools lost). Apart from the school data missing, pupil data are missing on variables such as FSM (*free school meal*), baseline scores on Mathematics and writing, age at KS1, and KS1 test scores. For more details about the data collection

procedures, see Hampshire County Council (1993). Matching of these test scores, other pupil data collected by schools and school level data, was carried out centrally by Hampshire LEA.

The main KS1 scores are those for the reading test, writing test, Mathematics test and Science assessment. For Science, we use the average score of teacher's assessments over four attainment targets, as there is no other test or task score available for this subject at Key Stage 1. The measures based on teachers' assessment for other subjects were not used.

The variables available from the baseline measurements at reception are *Average English score* (English baseline*), Average Mathematics score* (Maths baseline), *Average Science score* (Science baseline) and *Average total score* (Total baseline). The first three scores are correlated with the total score with correlation coefficients 0.89, 0.92 and 0.87 respectively. Only the first three were used for modelling their effects on each KS1 response. For all the pretest and posttest scores the scale scores have been transformed to Normality using Normal scores. This provides a common scale so that overall, each variable has a mean of zero and a standard deviation of 1.

The instrument for collecting baseline data is the individual pupil's Baseline Assessment Record.

Key stage 1 1992 to key stage 2 1996 (KS1KS2)

The data consists of 4,724 pupils from 114 schools, of which 25 schools are also involved in the baseline-KS1 study, although this information is not used in the present analyses. For language we use the scores of *Reading test* and *Writing test* to match the analysis in the early school period (Baseline to KS1). There are 4420 and 4392 records available for the two scores respectively with a correlation coefficient 0.89. The *main test score* is also used as it represents the overall score on English although it is highly correlated with the reading score ($r = 0.93$) and there are 4390 records available for this outcome.

For Mathematics, we use the *main test score* which has a high correlation with another four outcomes (*Test A score*, *Test B score*, *Main test level* and *Final test level*) with correlation coefficients of 0.96, 0.96, 0.92 and 0.92 respectively. This score again matches those in the baseline study and there are 4352 records available.

For Science, we use also the *main test score* with 4338 valid records.

Teacher assessments are not used at all.

Appendix 1 contains a list of all the variables, in addition to the baseline and test scores.

The next two sections describe, respectively, the fitting of multilevel models to the BKS1 data and the KS1KS2 data. This is followed by a description of how the results are interpreted and presented to schools.

Modelling Baseline to key stage 1 assessments.

Table 1 shows the results of fitting a basic model for each separate KS1 outcome. This model is simply a variance components model with a single intercept term and variances at pupil and school level. It may be written (Goldstein, 1995) as

$$y_{ij} = \beta_0 + u_{0j} + e_{0ij}$$
$$\text{var}(u_{0j}) = \sigma^2_{u0}, \qquad \text{var}(e_{0ij}) = \sigma^2_{e0} \tag{1}$$

where the u_{0j}, e_{0ij} respectively are the level 2 and level 1 'residuals'. All computations have been carried out using the MLn computer package (Woodhouse et al., 1996).

Table 1. Parameter estimates from fitting Model (1) (SE in brackets)

Parameter	***Reading***	***Writing***	***Mathematics***	***Science***
Intercept	-0.024(.027)	-0.008(.031)	-0.030(.029)	-0.034(.037)
School level variance	0.091(.013)	0.126(.017)	0.110(.015)	0.190(.024)
Pupil level variance	0.667(.012)	0.743(.013)	0.702(.013)	0.695(.012)
Intra-school correlation (%)	12.0	14.5	13.5	21.5

The Science intra-school correlation is fairly high, presumably because of the greater variation among teachers, since this is teacher assessed.

The next stage of analysis is to adjust for the baseline measures and for other relevant factors at either the pupil or school level. The first set of analyses uses the three baseline measures, their average values for each school and interactions among the measures. Table 2 gives the results of these analyses. The model being fitted can be written in general form, to include explanatory variables, as

$$y_{ij} = \sum_{h=0}^{p} \beta_h x_{hij} + u_{0j} + e_{0ij} \qquad (2)$$
$$\text{var}(u_{0j}) = \sigma^2_{u0}, \qquad \text{var}(e_{0ij}) = \sigma^2_{e0}$$

Table 2. Parameter estimates for baseline variables for KS1 outcomes.

	KS1 Reading		***KS1 Writing***		***KS1 Math***		***KS1 Science***	
Variable	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Fixed effect								
Intercept	-0.026	0.029	-0.020	0.034	-0.027	0.033	-0.128	0.068
English Baseline	0.149	0.014	0.169	0.015	0.116	0.014	0.141	0.015
Maths. Baseline	0.339	0.016	0.356	0.017	0.399	0.016	0.333	0.016
Science Baseline	0.080	0.016	0.085	0.017	0.067	0.017	0.112	0.017
English/Maths interaction	0.009	0.009	0.026	0.010	0.009	0.009	-0.040	0.015
School Mean English	-0.353	0.052	-0.360	0.061	-0.350	0.060	-0.501	0.123
(School Mean English)^2	-0.005	0.070	-0.014	0.081	-0.037	0.079	0.493	0.335
English/Science interaction							0.061	0.016
Variances								
School level	0.073	0.010	0.106	0.014	0.102	0.013	0.169	0.021
Pupil level	0.494	0.009	0.547	0.010	0.510	0.009	0.517	0.009
Intra-school correlation (%)	13.9		16.2		16.7		24.6	

Table 3 shows the results of the variance components analysis using all the explanatory variables available on each student.

Table 3. Parameter estimates for all explanatory variables* for KS1 outcomes. (parameter estimates not significant at the 5% level are omitted from the analysis). Standard errors in brackets.

Variable	KS1 Reading	KS1 Writing	KS1 Mathematics	KS1 Sciences
Baseline variables				
English baseline	0.089(0.013)	0.099(0.014)	0.079(0.013)	0.108(0.014)
Maths baseline	0.225(0.014)	0.237(0.015)	0.278(0.015)	0.226(0.015)
Science baseline	0.046(0.014)	0.046(0.015)	0.030(0.015)	0.079(0.016)
English/Maths interaction	0.042(0.008)	0.055(0.009)	0.041(0.009)	-.006(0.013)
English/Science interaction	-	-	-	0.056(0.015)
School mean English baseline	-.281(0.039)	-.276(0.050)	-.270(0.052)	-.326(0.067)
(School mean English baseline)^2	-.112(0.051)	-.136(0.064)	-.121(0.068)	-
Other pupil variables				
SENAUDIT	-.449(0.012)	-.438(0.013)	-.422(0.013)	-.376(0.013)
KS1AGDYS	0.024(0.030)	0.129(0.032)	0.251(0.031)	0.188(0.031)
Girl	0.107(0.016)	0.171(0.017)	-.211(0.017)	-.217(0.017)
FSM	-.083(0.023)	-.161(0.025)	-.081(0.024)	-.143(0.025)
Absence	-.003(0.0005)	-.003(0.0006)	-.002(0.0006)	-.003(0.0005)
TTS	0.010(0.019)	0.008(0.021)	0.018(0.022)	0.021(0.024)
TTS^2	-.008(0.004)	-.012(0.004)	-.010(0.004)	-.006(0.004)
TOS	0.035(0.024)	0.049(0.026)	0.056(0.026)	0.045(0.027)
Other school variables				
SFSMTPC	-.008(0.001)	-.007(0.002)	-.008(0.002)	-.010(0.002)
VERGRP	0.102(0.047)	0.076(0.054)	0.048(0.054)	0.199(0.059)
SAUDY2PC	-	-	0.005(0.002)	-
PREVSCPC	-.601(0.288)	-1.095(0.360)	-.535(0.375)	0.502(0.460)
Controlled school	-	-	-	-.227(0.083)
Aided school	-	-	-	-.229(0.100)
NOEMPAD	0.065(0.031)	0.059(0.038)	0.095(0.039)	0.240(0.047)
School level variance	0.033(0.005)	0.061(0.008)	0.071(0.010)	0.135(0.017)
Pupil level variance	0.393(0.007)	0.439(0.008)	0.421(0.008)	0.438(0.008)
Intra-school correlation (%)	7.7	12.2	14.4	23.5

* The variables: number of teachers and class size were not significant for any of the outcomes.

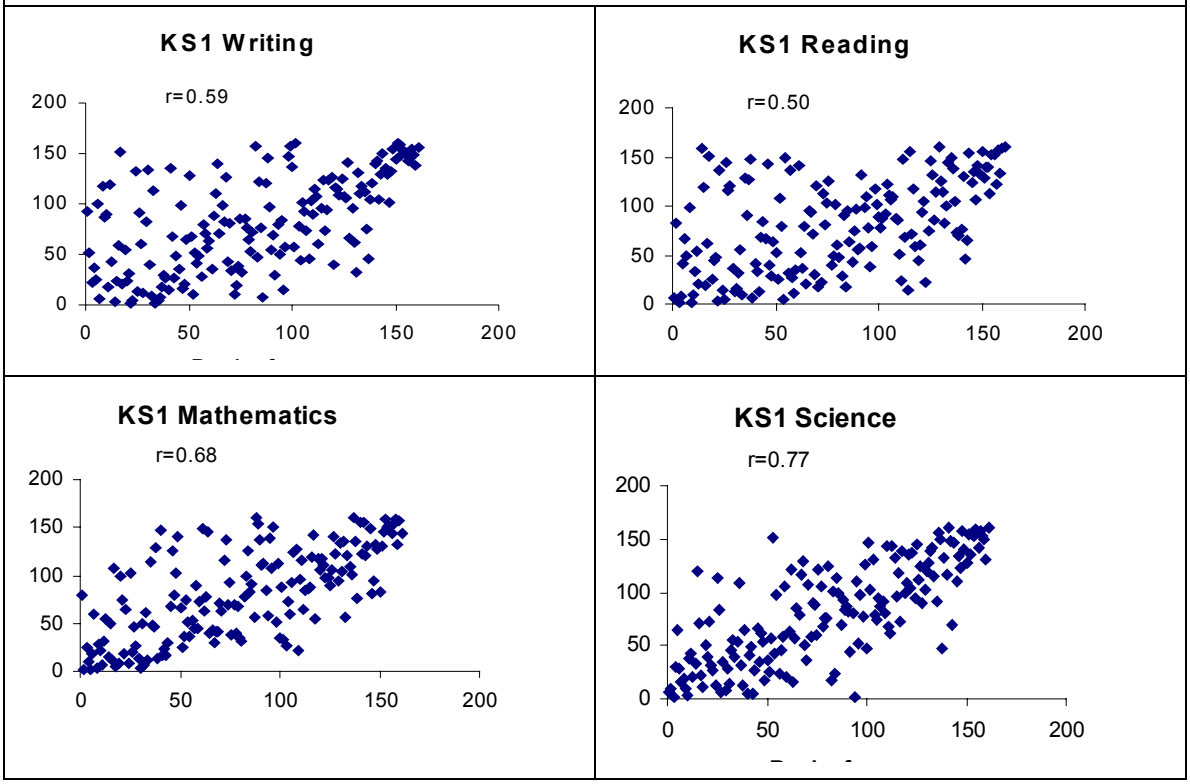
The intra-school correlations are somewhat lower here except for Science, which also has the highest value of any subject. This reflects the fact that it was teacher assessed and incorporates the additional between-teacher variation that this implies.

Adjusted, value added estimates

From the results in Table 3 we can compute estimates of the school residuals, u_{0j}, in (2) and these are the adjusted or 'value added' estimates for each school. Associated with each of these is a standard error which measures the uncertainty associated with sampling variation, and we shall make use of this information below. We note that these residuals are 'shrunken', so that their variance is less than the true between-school variance as estimated in Table 3. The fewer the number of students in a school the greater the shrinkage towards zero, so these estimates may be regarded as 'conservative' in that they give less weight to schools with small numbers which will tend to be subject to large sampling fluctuations.

To illustrate the effect of adjusting for the baseline and other pupil level variables, In figure 1 we show the relationship between the ranks of the raw mean scores for each outcome and the adjusted residuals from the analysis in Table 3.

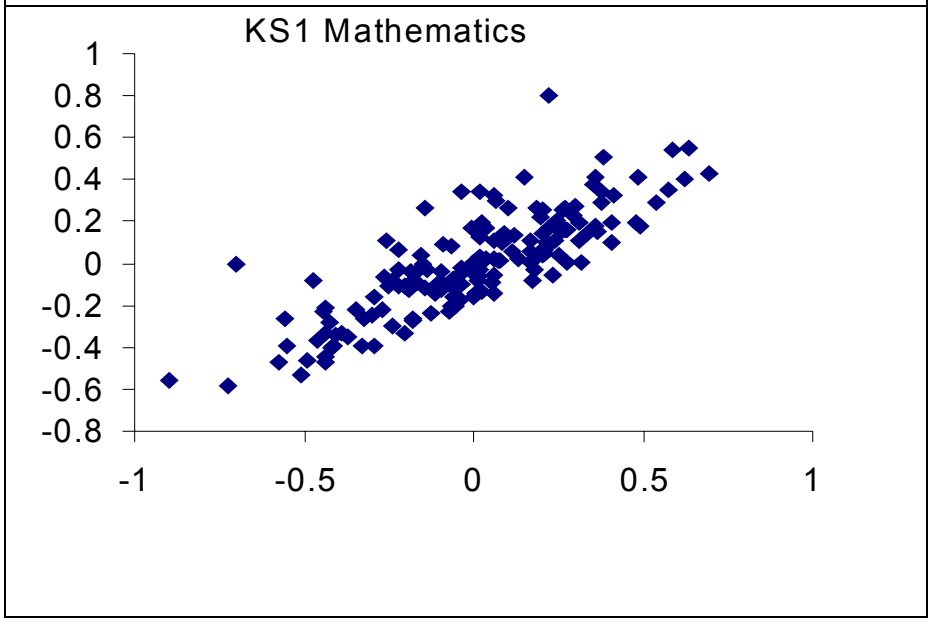
Figure 1. Adjusted school level residuals (Y axis) by school raw means (X axis) for KS1 outcomes. The scale values are ranks.



The smallest correlation is for reading and all the plots show how misleading it can be to judge a school by its raw mean scores.

Figure 2 shows a plot of the residuals, for Mathematics, estimated from Tables 2 and 3, to illustrate that the addition of the further pupil and school variables makes some difference when adjusting for school differences. The differences are somewhat more marked when adjusting for reading and writing.

Figure 2. School level residuals, for Mathematics, adjusted for all pupil and school level variables (Y axis) against residuals adjusted for baseline variables only (X axis). Correlation = 0.72



Differential effectiveness

The variance components models so far considered assume that there are only simple differences between schools. The research evidence on school effectiveness, however, (Goldstein, 1997, Thomas et al, 1997) shows that schools differ along a number of dimensions, and in particular that the coefficient of the intake score varies between schools. We have therefore explored a number of models in which the model coefficients are assumed random at the school level. For each outcome, one or more of the baseline variables are found to have significant random coefficients. There is also evidence that

the gender difference, the SEN effect and the age of the pupil relationship vary across schools. For present purposes, however, we have used only the baseline random coefficients.

To simplify the analysis and presentation still further, we have chosen, for each outcome, the linear combination of baseline scores estimated and presented in Table 3 to define a composite variable which is entered as a single variable into the fixed part of the model with a random coefficient. In each case the resulting model fits almost as well as the full model with one or more separate random coefficients.

Table 4 presents the final analyses using these composite variables, which we denote by 'English-composite', 'writing-composite', etc., and includes also a random coefficient for the composite at the school level.

Table 4. Parameter estimates for the model of Table 3 with baseline scores replaced by a single composite for each outcome and a random coefficient of the composite at level 2. Standard errors in brackets.

Variable	KS1 Reading	KS1 Writing	KS1 Mathematics	KS1 Sciences
Baseline composite	1.018(0.046)	1.017(0.049)	1.014(0.044)	1.009(0.056)
Other pupil variables				
SENAUDIT	-.458(0.012)	-.447(0.013)	-.429(0.013)	-.389(0.013)
KS1AGDYS	0.032(0.030)	0.125(0.032)	0.254(0.031)	0.192(0.032)
Girl	0.105(0.016)	0.168(0.017)	-.212(0.017)	-.218(0.017)
FSM	-.084(0.023)	-.163(0.025)	-.079(0.024)	-.145(0.024)
Absence	-.003(.0005)	-.003(.0005)	-.002(.0005)	-.003(.0005)
TTS	0.010(0.019)	0.009(0.021)	0.020(0.022)	0.024(0.023)
TTS^2	-.008(0.004)	-.012(0.002)	-.010(0.004)	-.010(0.002)
TOS	0.033(0.024)	0.048(0.026)	0.057(0.026)	0.042(0.027)
Other school variables				
SFSMTPC	-.008(0.001)	-.007(0.002)	-.008(0.002)	-.010(0.002)
VERGRP	0.096(0.048)	0.072(0.053)	0.056(0.054)	0.205(0.059)
SAUDY2PC	-	-	0.005(0.002)	-
PREVSCPC	-.463(0.294)	-.984(0.353)	-.483(0.374)	0.542(0.459)
Controlled school	-	-	-	-.220(0.083)
Aided School	-	-	-	-.221(0.101)
NOEMPAD	0.065(0.031)	0.056(0.037)	0.095(0.039)	0.241(0.046)
School level variance				
Var(intercept)	0.034(0.005)	0.061(0.008)	0.072(0.010)	0.137(0.017)
Cov(intercept, composite)	0.004(0.009)	0.027(0.013)	0.007(0.012)	0.034(0.021)
Var(composite)	0.122(0.031)	0.161(0.036)	0.107(0.027)	0.278(0.048)
Pupil level variance	0.385(0.007)	0.427(0.008)	0.412(0.007)	0.414(0.008)

We shall postpone to a later section a discussion of the presentational issues.

Socio economic group

Information on the socio economic group of 20% of the students was available. While this was not a representative sample of the students, it is of some interest to see if the inclusion of this variable would alter any of the inferences, especially whether it would alter appreciably the school rankings.

In fact the rankings are hardly altered, nor are the school and pupil level variance estimates in a variance components model, for each outcome. The inclusion of other variables, in addition to baseline, such as SEN stage, absences and free school meals eligibility adjusts for most of the socio economic group effect. In the remaining analyses we therefore shall omit socio economic group.

KS1 to KS2 analyses

In this analysis the analysis is based upon approximately 4400 students with scores at KS1 and KS2 in 114 schools. The same set of predictor variables as for the earlier analyses are used. In 1992, the first year of statuary tests and tasks for Key Stage 1, there were tests or tasks in Reading, Writing, Spelling and Handwriting, Number and Shape and Space, and Science. These were graded by levels on each test or task. The analysis uses the following KS1 scores:

- English task score - a combination of four component test scores
- Mathematics task score - a combination of two component scores
- Science task score - a single score

The outcome variables we have chosen are as for the earlier analysis, namely Reading, Writing, Mathematics, and Science.

A sequence of similar analyses to the earlier ones was carried out. Tables 5-8 parallel Tables 1-4 for the earlier analyses. Both the response variables and the KS1 predictors are standardised: the response variables have distributions very close to Normality.

The results for KS2 outcomes are similar to those from the earlier analyses, with a few differences, but we shall not comment on these in detail. One feature is worth noting, namely that the addition of variables other than the KS1 test scores does not explain as much further between-school variation as in the earlier analyses. Figure 3 illustrates the effect of the addition of the further pupil variables which is less important than is the case for the earlier analyses, and likewise for the Reading and Writing outcomes.

Figure 3. School level residuals, for Mathematics, adjusted for all pupil level variables (Y axis) against residuals adjusted for KS1 variables only (X axis).

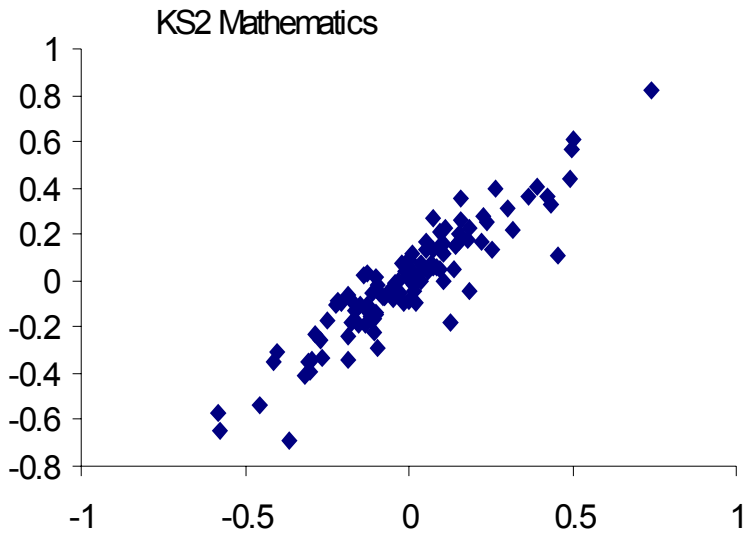


Table 5. Parameter estimates from fitting Model (1) for KS2 outcomes (SE in brackets).

KS1 score	Intercept	School variance	Pupil variance	Intra-school correlation (%)
Reading	0.022(0.041)	0.142(0.024)	0.874(0.019)	14.0
Writing	0.014(0.045)	0.180(0.029)	0.818(0.018)	18.0
Mathematics	0.016(0.043)	0.157(0.026)	0.853(0.019)	15.5
Science	0.021(0.051)	0.248(0.038)	0.774(0.017)	24.3

Table 6. Parameter estimates for KS1 variables for KS2 outcomes. Standard errors in brackets.				
KS1 Variable	Reading	Writing	Mathematics	Science
Intercept	-0.020(0.026)	-0.024(0.036)	-0.024(0.031)	-0.028(0.039)
English	0.504(0.017)	0.509(0.017)	0.297(0.016)	0.336(0.017)
English school . mean	0.284(0.081)	0.246(0.108)	0.407(0.094)	0.504(0.115)
Mathematics	0.158(0.018)	0.081(0.018)	0.409(0.017)	0.240(0.017)
Mathematics school mean	-0.186(0.094)	-0.277(0.126)	-0.352(0.109)	-0.315(0.135)
Science	0.110(0.019)	0.059(0.020)	0.126(0.019)	0.174(0.019)
Science school mean	-0.268(0.091)	-0.096(0.121)	-0.296(0.106)	-0.227(0.131)
School Variance	0.050(0.009)	0.108(0.018)	0.079(0.013)	0.133(0.021)
Pupil Variance	0.530(0.012)	0.546(0.012)	0.484(0.011)	0.490(0.011)
Intra-school correlation (%)	8.6	16.5	14.0	21.3

Table 7. Parameter estimates for all explanatory variables* for KS2 outcomes. (parameter estimates not significant at the 5% level are omitted from the analysis). Standard errors in brackets.

KS1 Variable	Reading	Writing	Mathematics	Science
Intercept	0.007(.026)	0.039(.043)	0.160(.034)	0.152(.037)
KS1 variable				
English	0.381(.018)	0.384(.018)	0.278(.017)	0.310(.018)
English school mean	0.164(.082)	0.300(.106)	0.220(.0101)	0.403(.0105)
Mathematics	0.159(.017)	0.090(.017)	0.376(.017)	0.211(.017)
Mathematics school mean	-0.244(.084)	-0.309(.120)	-0.351(.0102)	-0.375(.0124)
Science	0.104(.019)	0.057(.019)	0.110(.018)	0.162(.018)
Science school mean	-0.187(.080)	-0.090(.119)	-0.239(.099)	-0.158(.0118)
Pupil variable				
Girl - Boy	0.169(.023)	0.237(.023)	-0.261(.022)	-0.187(.022)
Free meal - None		-0.067(.034)		-0.092(.033)
Special needs stage	-0.282(.017)	-0.232(.017)	-0.216(.016)	-0.182(.017)
Sessions absent	-0.002(.0006)	-0.002(.0007)	-0.002(.0007)	-0.002(.0007)
Terms in this school			0.018(.006)	
Age in months			-0.010(.003)	-0.007(.003)
School/class variable				
No. teachers in class (-1)		-0.271(0.104)		
% of pupils at Step 0	-0.004(.002)			
% Entitled to FSM	-0.007(.003)		-0.008(.003)	
% Absence				-0.052(.013)
Controlled - County		-0.254(.090)		
School Variance	0.034(.007)	0.097(.016)	0.067(.012)	0.103(.0167)
Pupil Variance	0.486(.011)	0.503(.011)	0.446(.010)	0.466(.010)
Intra-school correlation (%)	6.5	16.2	13.1	18.1

* The variables: number of teachers and class size were not significant for any of the outcomes. The KS1 test scores, number of sessions absent and age are centered close to their means

As in the earlier analyses a composite variable of the KS1 scores has been used for the final analysis from which the residual estimates are derived. These results are given in Table 8.

Table 8. Parameter estimates for the model of Table 7 with KS1 scores replaced by a single composite for each outcome and a random coefficient of the composite at level 2. Standard errors in brackets.

Variable	KS2 Reading	KS2 Writing	KS2 Mathematics	KS2 Sciences
KS1 composite (*Z*)	1.016(0.033)	1.005(0.039)	1.018(0.032)	1.016(0.036)
Other pupil variables				
SENAUDIT	-.284(0.017)	-.230(0.017)	-.214(0.015)	-.175(0.016)
KS2AGDYS	-	-	-.113(0.037)	-.092(0.038)
Girl	0.163(0.022)	0.234(0.023)	-.262(0.021)	-.189(0.021)
FSM	-	-.071(0.034)	-	-.077(0.032)
Absence	-.002(.0006)	-.002(.0007)	-.002(.0006)	-.003(.0007)
TTS	-	-	0.014(0.006)	-.009(0.006)
Other school variables				
SFSMTPC	-.006(0.002)	-	-.008(0.002)	-
NOTEACH	-	-.290(0.105)	-	-
%SENAUDIT	-.004(0.002)	-	-	-
%Absence	-	-	-	-.054(0.012)
Controlled school	-	-.185(0.095)	-	
School level variance				
Var(intercept)	0.033(0.007)	0.100(0.017)	0.064(0.011)	0.098(0.016)
Cov(intercept, Z)	0.002(0.006)	.0007(0.012)	0.014(0.008)	0.009(0.011)
Var(Z)	0.034(0.012)	0.051(0.017)	0.047(0.012)	0.052(0.014)
Pupil level variance	0.476(0.011)	0.493(0.011)	0.426(0.009)	0.451(0.010)

The same broad picture emerges in terms of the factors associated with progress between Key Stages as in the Baseline to Key Stage 1 analyses. These are gender, with girls making more progress in Reading and Writing and less in Maths and Science; SEN stage, and free school meals, at least in Writing and Science. Other factors appear of lesser importance.

Presenting the results to schools

A major aim of the project is to provide individual schools with data which will allow them to make valid comparisons of their performance compared with the other schools in the LEA. The criteria used are as follows.

1. The information provided to each school is confidential to that school and the LEA. Summary information about the performance of all LEA schools is provided solely for comparative purposes.
2. The system is seen as evolving over time in the light of feedback from users, availability of new data and national developments.
3. It is recognised that value added estimates have uncertainty attached to them and that this needs to be taken account of in any interpretation.

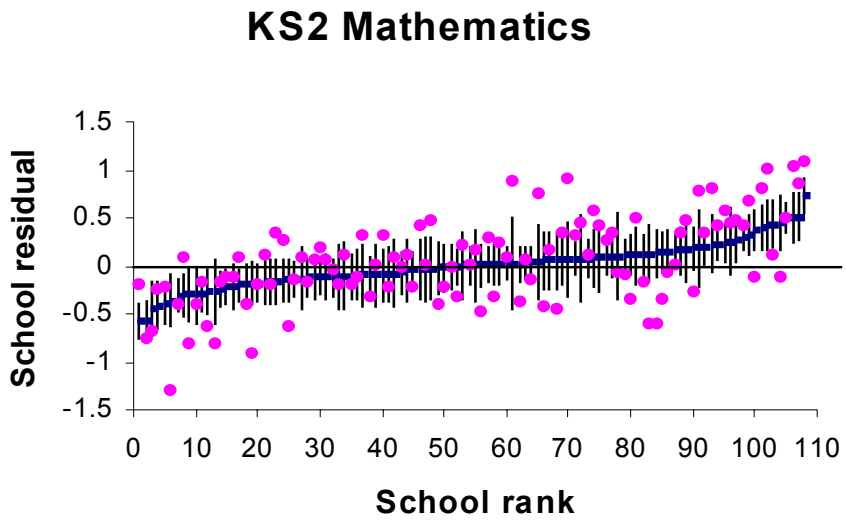
For each outcome there are two 'residual' estimates; an intercept and a 'slope' derived from an analysis such as those in Tables 4 and 8. We shall discuss presentation using the KS2 Mathematics outcome: similar issues apply to the other outcomes at KS1 and KS2.

Value added rankings and uncertainty intervals.

We first present the results of fitting the variance components model of Table 7 where a single residual is estimated for each school. These residuals are ordered and for each one a 95% confidence or uncertainty interval is given. As can be seen clearly, only a minority of schools are significantly different from the overall mean of zero. An adaptation of this graph allows for the comparison of pairs of schools. If the confidence intervals are scaled down by a factor of approximately 0.7 (see Goldstein and Healy, 1995), we may judge two schools as significantly different if and only if their intervals do not overlap. Typically, only about a third of all such comparisons yield significant results. While this finding is not especially relevant to the use of the results by schools as described above, it does imply that the use of such data for the *public* ranking of schools in 'league tables' is problematical.

Figure 4 shows these residuals for KS2 Mathematics, together with the raw mean scores for each school. It is clear that the ordering by value added scores is quite different from that implied by the unadjusted mean scores.

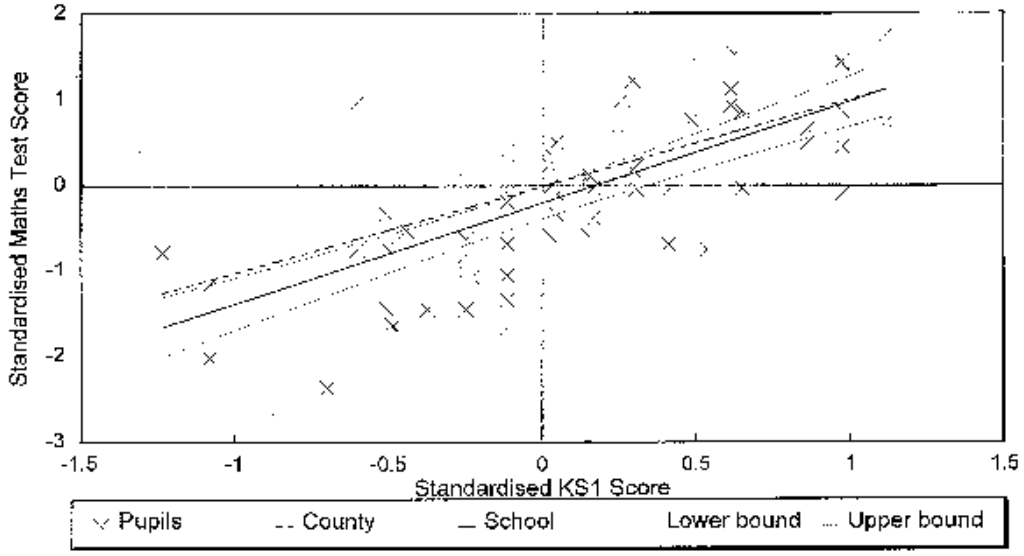
Figure 4. KS2 Mathematics residual estimates, 95% confidence intervals and unadjusted mean scores.



Differential effectiveness plots

The above result provides only a simple average estimate of 'value added' for schools and is of limited usefulness to each school. Based on the results of the analyses in Table 8 we can estimate residuals associated with the 'intercept' and the coefficient of the composite KS1 score and so determine a line for each school. Figure 5 gives an example for a school.

Figure 5. A differential value added graph for one school for Mathematics outcome at KS2.



The County (LEA) line is drawn to pass through the origin and each school's line is drawn in relation to this using the residual estimates. For individual pupils the raw or total residual from the fixed part of the model is used and added to the County line at the appropriate KS1 composite score for that student.

The (continuous) school line here has a higher estimated slope than that for the LEA (County) but is below the latter except for the higher achieving students at KS1. The 95% confidence interval around the school line (indicated by the Lower and Upper bounds) suggests that the school line is *significantly* below the LEA line for all students who have KS1 scores below average. For the school this suggests that they should pay attention to their arrangements for such students to see if there is an explanation and, if it is accepted that there is a real problem, whether it can be remedied. The individual pupil scores are also a useful diagnostic since they will be identified to the school and studied to see if there are particular explanations associated with 'outliers' or possibly students with particular characteristics which have not been incorporated into the model as adjustment factors.

Each school will have one of these graphs available for each subject at each key stage and will therefore be able to compare performances. A programme of in service training has been implemented, initially for head teachers and advisors, to familiarise users with

interpretations, including the limitations of these results. A programme to evaluate the use of these results is also being set up.

Hampshire and the Institute of Education plan to continue the analysis of baseline to KS1 results in future years. The absence of sufficient KS1 data for 1993 and 1994 means that it is not possible to return to the analysis of KS1 to KS2 data until 1999.

Discussion

The Hampshire value added project has demonstrated that it is both feasible and useful to provide sensitive analyses of students progress which yield important school improvement data for use by primary schools. It has shown that complex multilevel models can be presented in ways that are accessible to potential users without sacrificing the essential components which emerge from the analyses.

The first results were available in July 1997 and it is intended that they will be replicated each year as data allow. Furthermore, over time it will become possible for schools to study their individual change in value added measures, so increasing the usefulness of the data. A programme of in service training has been set up as has a programme to evaluate the use of the information. In addition, a study is being made of how governors and parents in particular can be brought into discussions of results.

A key feature of the analysis and presentation is the emphasis on uncertainty surrounding the results. The provision of uncertainty bands appears to pose no real interpretational difficulties for users. In addition, however, there are a number of other limitations of the analyses which need to be understood by users.

The first limitation is that the residual estimates are derived from a model which uses only the data available for analysis. It is possible that there are other factors which ought to be included but for which data are not available. One such factor, socio-economic group, has been eliminated as a source of serious misspecification. Another area of concern, especially for the KS1 analyses, is the quality and reliability of the baseline measures. It is known (Woodhouse et al., 1996) that predictor variables with low reliability can affect residual and parameter estimates, and this is an issue which will be addressed in future research. This is being addressed in the new Baseline materials

developed by Hampshire LEA. These have been accredited by the Qualifications and Curriculum Authority (QCA) and piloted in 1997/98. A further factor is that many students change school during a particular phase of education. Between Baseline entry and KS1 tests just over 5% had changed schools at least once and between the KS1 tests and KS2 tests 37% had changed, with 4% changing more than once. The analyses show that the number of previous schools and the length of time in the final school are associated with progress. In principle the analysis could incorporate information about all the schools attended and apportion a students progress among them (Hill and Goldstein, 1998). It is difficult to acquire such data, however, although further work into the feasibility of doing this is being carried out.

All these limitations about the adequacy of the model add a further dimension of uncertainty and need to be borne in mind when making interpretations. In our view the strength of the results is that they provide a further, quantitative, indicator that schools can use in their judgements of how well they are functioning. The results are in the nature of screening instruments that can indicate where problems may be present, but which are not precise diagnoses. In particular their use as public accountability measures, e.g. in the form of performance tables or 'value added league tables' is inappropriate and would destroy their credibility and usefulness. If they were ever to become 'high stakes' pieces of information like the current DfEE league tables of examination results, then they would inevitably become distorted and no longer reflect any underlying reality of school performance. The same conclusions, of course, apply to any future national scheme which attempts to derive value added measures.

References

DfEE (1997). *Excellence in Schools (White Paper)*. London, Department for Education and Employment.

Goldstein, H. (1995). *Multilevel Statistical Models*. London, Edward Arnold: New York, Wiley.

Goldstein, H. and Healy, M. J. R. (1995). The graphical presentation of a collection of means. *J. Royal Statistical Society, A.* **158**: 175-7.

Goldstein, H. and Spiegelhalter, D. J. (1996). League tables and their limitations: statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society, A.* **159**: 385-443.

Hampshire, (1993). *Baseline Assessment of Achievement, Teachers' Resource Booklet*: Hampshire Assessment and Recording of Achievement Team; Winchester, Hampshire County Council.

Hill, P. W. and Goldstein, H. (1998). Multilevel modelling of educational data with cross classification and missing identification of units. *Journal of Educational and Behavioural statistics* **23**: 117-128.

Woodhouse, G. 1996*: Multilevel Modelling Applications, A Guide for Users of MLn*. Multilevel Models Project, Institute of Education, University of London.

Woodhouse, G., Yang, M., Goldstein, H. and Rasbash, J. (1996). Adjusting for measurement error in multilevel analysis. *Journal of the Royal Statistical Society, A.* **159**: 201-12.

Appendix 1. Variables used in the study, in addition to baseline and test scores.

Pupil level variables

Gender	0=Boy, 1=Girl
FSM	Free school meal, 0=not entitled, 1=entitled
SEN	Special education need statement, 0=no, 1=yes
SENAUDIT	Stage of special education need, 0,1,2,3 stage
NURS	LEA nursery education, 0=no, 1=yes
TTS	No. of terms in this school
TOS	No. of terms in other schools
NOSCHS	No. of schools attended
Absence	No. of half days absent
E2L	English as a 2nd language, 0=no, 1=yes
BirthMTH	Month of birth, 1=Sept., 2=Oct. …….
BLAGEDYS	age in days at KS1 completion
KS1AGEDYS	Age in days at KS1 completion (15/5/96), divided by 365 then centred at 7
SEG (*a voluntary return)*	Social economic group of parents, 1=manual, 2=intermediate, 3=professional

School level variables

VERTGR	Vertical grouping 0=0-1 groups, 1=2+ groups
NODIFF	No. different teachers for this class, recoded as 1=0-1, 2=2-5, 3=team teaching
NOPUPILS	No. pupils in this class, recoded as 1 for ≤ 20, 2 for 21-25, 3 for 26-30, 4 for ≥ 31
NOTEACH	No. of teachers in this class at Form 7 date, recoded as 0=3-10, 1=0-2
NOEMPADS	No. other employed adults in this class
STOTNOR	School –total No. on Roll, centred at 250
SDENOM	School denomination, 1=country, 2=controlled, 3=aided
STYPE	School type, 1=infant, 2=first/infant & 1st year, 3=junior & infant, 4=junior

SY2NOR	No. on roll at Year 2, centred at 60
SFSMTPC	School percentage entitled to FSM, centred at 20%
SFSMY2PC	School percentage entitled to FSM at Year 2, centred at 20%
SSENTPC	School percentage pupils with Statements
SAUDITPC	School percentage pupils on SEN register, centred at 20%
SAUDIY2PC	School percentage pupils on SEN register at Year 2, centred at 25%
SABSTPC	School percentage half day absence, centred at 5%
SPECUNIT	School with special unit, 0=no, 1=yes
NURSUNIT	School with nursery unit, 0=no, 1=yes
PREVSCPC	Proportion of pupils changed schools previously, 0.0 – 0.33,
	(No. of pupil changed schools)/ (Tot. no. pupils in this school)