The use of value added information in judging school performance

by

Harvey Goldstein

Institute of Education, University of London

Pan Huiqi

Institute of Child Health, University College London

Terry Rath and Nigel Hill

Hampshire Local Education Authority

Address for correspondence:

Professor H. Goldstein

Institute of Education, London, WC1H 0AL.

h.goldstein@ioe.ac.uk

Summary

It is now generally recognised that in order to obtain fair comparisons of achievement standards among schools, the minimum requirement is a ‘value added’ analysis which adjusts for intake achievements. It has been suggested, however, that it may be possible to *approximate* full value added analyses of Key Stage test score data using certain kinds of cross sectional data which are readily available, such as free school meal eligibility and average Key Stage test scores for younger cohorts of children. This issue has been of particular interest to OFSTED because of the kinds of information available to inspection teams and OFSTED has provided financial support for the present study.

Key Stage 2 (KS2) data from Hampshire primary schools have been used to study the properties of such so called ‘proxy’ measures for use as performance indicators, in comparison with full value added measures. The proxy data involve measurements taken at the same time as the KS2 tests, such as free school meal eligibility and average KS1 results for a younger cohort. The research reported here shows that such proxy measures are inadequate and likely to produce misleading comparisons among schools, providing little improvement over the use of completely unadjusted (‘raw’) KS2 scores. These results therefore provide little support for the routine use of such proxy measures either for inspection purposes or for publishing as ‘league table’ rankings of schools. They do not constitute valid measures of educational quality. If, despite their misleading nature, either unadjusted or proxy measures are presented, their nature and limitations should be fully documented and understood by those who would wish to use them. The results of the present research provide a basis for doing this.

The results reported here also raise wider questions about how detailed value added analyses, if and when these become available on a large scale, can be used for school improvement purposes by the use of informed feedback to schools. A crucial issue will be how the results can be presented in ways that avoid the well known disadvantages associated with any kind of public rankings (league tables).

Acknowledgements

We are grateful to OFSTED for funding this research and to Chris Bryant, Christine Agambar and Audrey Brown for helpful advice. The views expressed are those of the authors.

1. Introduction - the Hampshire value added project

Hampshire LEA have been developing a system of value added analysis and reporting for primary schools since 1996. In this scheme linked baseline to KS1 and KS1 to KS2 data are obtained for all schools, a full value added analysis carried out and the results fed back to schools in the form of charts, by the separate subjects of English, mathematics and Science, which provide estimates of the school's position in comparison with the LEA as a whole. There are three key features of the scheme.

The first is that the results are presented with confidence intervals which show a range for each school so that differences can be properly contextualised. The second feature is that a 'differential effectiveness' model is used which allows judgements to be made taking into account the KS1 achievement. Thus, for example, a school may show average progress for high achieving KS1 children but low progress for low achieving KS1 children; this kind of differential 'screening' is found to be particularly useful in identifying potential strengths and weaknesses. The third feature is that the value added scores remain confidential to the school and the LEA and there is an undertaking that they will not be used for the purposes of producing public 'league tables'. This last feature is considered by the participants to be especially important for the success of the scheme since it discourages any attempt to 'play the system' as almost certainly occurs with published 'performance indicators' based upon 'raw' or unadjusted test scores. While the scheme was being developed participation was voluntary; the Hampshire Education committee now expect all schools with pupils between KS1 and KS2 to take part.

The data used in the present study are the Primary schools taking part in the Hampshire value added project (Yang et al., 1999) who had matched data on the same children at KS2 in 1998 and KS1 in 1994. There were 76 schools with 1400 children in total who had data on the variables listed in Table 1. It was not possible to obtain matched records for the junior schools. The restricted nature of this sample suggests that some caution should be exercised when extrapolating to all schools, but there is no reason to think that the general pattern of results would change.

2. OFSTED inspections

When OFSTED inspection teams prepare for inspections they are given a great deal of information about the performance of pupils on key stage tests and examinations. For primary schools, for example, they will have available the average scores in Maths, English and Science for the most recent and preceding cohorts at KS1 and KS2. Inspectors are expected to use such information, together with other 'contextual' factors such as the proportion of pupils eligible for free meals, to make preliminary judgements about the 'performance' of the school.

The problems with using such information are now well known and have been studied in many research contexts; for a non-technical account see Goldstein (1997). The use of value added data which makes adjustments for initial achievement is now widely accepted by government (see DfEE, 1995) and others. In practice, however, value added analysis is only possible where linked individual student data are available, for example between KS1 and KS2, and this is only the case in a few LEA areas. This raises the question as to whether there is any way of approximating to a full value added analysis using data which are widely available and the purpose of the present research is to study this possibility.

It has been suggested that the use of aggregate school level performance data, for different cohorts of children, might be an acceptable substitute, or at least provide more valid information about schools' effectiveness than the use of unadjusted key stage test scores or test scores adjusted only for aggregate level characteristics such as the percentage of children in the school eligible for free school meals. Few LEAs and schools currently have linked information across key stages on individual pupils to carry out analyses at the level of detail in Hampshire. The following analyses were therefore carried out to investigate the usefulness of these measures in the context of school inspections, using the Hampshire data to make comparisons among methods. The research was funded by OFSTED and carried out by the authors.

The analyses explore this issue by comparing various 'proxy' models with unadjusted mean scores and with a full value added analysis in order to see the difference that this makes to judgements about individual schools. The measures used are those given in Table 1. School denomination was also studied but did not contribute significantly to the predictions.

Several exploratory analyses were carried out and the main findings are reported below.

The response variables used in this analysis, reading and Mathematics test scores, have been transformed to have a standard Normal distribution. In order to assist in the interpretation of the parameters in the various models fitted we note that on this scale the age coefficient is 0.33 for English and 0.25 for Mathematics. These represent the average differences in scores for children one year apart in age, taking the test at the same time. This difference is not equivalent to the average change in test score made during the course of a year as a child gets older. Since these tests are given only once to a cohort of children, which is also the case for the key stage tests, we cannot directly estimate the age coefficient associated with progress through time. Goldstein and Fogelman (1974), however, estimate the ratio of this coefficient to the 'contemporaneous' coefficient to be 2.6 which gives estimates for the former of 0.9 for English and 0.7 for Mathematics. These values may be used, with care, to form approximate judgements.

1.1 Table 1. Measures used in various models	
Variable	**Comments**
gender	0=boy, 1=girls
% girls in year 6	Computed from individual level data.
FSM	0= not eligible, 1= eligible for free school meals
%FSM	% eligible in school *
Age	Age of child in years at time of KS2 test measured about 11.0 years
KS2 Mathematics	KS2 overall Mathematics test score for child in 1998, transformed to N(0,1) scores.
KS2 English	KS2 overall English test score for child in 1998, transformed to N(0,1) scores.
KS1 English	The mean KS1 English score for 1994 based upon the four English component tasks (level data) for a child**
KS1 Mathematics	The KS1 Mathematics score for 1994 based upon level data for a child**
KS1 composite	A composite KS1 score for 1994 based upon KS1 English and Mathematics, for Mathematics and English responses***
Average KS1 1998	School mean test score for 1998 KS1 children in reading, writing and Mathematics****
Average KS1 1997	School mean test score for 1997 KS1 children in reading, writing and Mathematics****
Average KS1 1996	School mean test score for 1996 KS1 children in reading, writing and Mathematics****
Average KS1 1995	School mean test score for 1995 KS1 children in reading, writing and Mathematics****
* The OFSTED categorisation (0-8, -20,-35, -50, 50+) yields only 6% in the third category and none in the fourth and fifth. This compares with 33% in the final three categories nationally (DfEE, 1998). The mean is 7.7%. The relationship with both outcomes is effectively linear in the range 0-30%.	
** The OFSTED scoring system for levels was used here. (level 1=1.5; level 2C=2.17; level 2B=2.5; level 2A=2.83; level 3=3.5; level 4=4.5)	
*** The following composite scores are used based upon preliminary analyses to establish the coefficient values: English: $1.132x_1 + 0.183x_2$. Mathematics: $0.647x_1 + 0.602x_2$. $x_1 = KS1$ English $x_2 = KS1$ Maths	

**** For technical reasons the scoring system for KS1 scores used here was as follows:

Level 1=1, level 2C=2, level 2B=2.33, level 2A=2.67, level 3C=3, level 3B=3.33, level 3A=3.67, level 4C=4, level 4B=4.33, level 4A=4.67. The correlation between these scores and the OFSTED scoring system is 0.98 at the pupil level and 0.99 at the school level. This implies that inferences are robust to the particular scoring system used.

3. Analyses

The full details of the analyses are given in tables in Appendices A and B: here we summarise the main results. Multilevel analyses are carried out throughout in order to take account of the between and within school variation in test scores. The analyses all use a two level model and have been carried out using the software package *MLwiN* (Rasbash et al., 1999)

The first set of analyses looks at using just the current (1998) KS2 data to adjust the outcome scores. Then the average KS1 scores for the school are added for the years 1998 through to 1995 (the appropriate 'cohort' for the 1998 KS2 pupils is in fact 1994 but these data were only available for a subsample). Following this a full value added analysis is done using the KS1 data matched for the same KS2 pupils. The basic findings are as follows, first for English.

The percentage of pupils eligible for free school meals (a school level variable) remains as a significant predictor throughout the analyses of the English score, with those schools having the lowest percentage having a predicted score up to about 0.25 more than the average. Pupil eligibility (i.e. measured at the pupil level) for free school meals is associated with a score reduction up to about 0.4 compared to not, as does being a boy rather than a girl. The average for the KS1 reading scores is the only useful aggregate level achievement predictor, with the 1995 average providing the best predictor. The 1998 average on its own also provides a weak prediction, but adds little when the 1995 average is fitted. Restricting attention to those data for 1998 only, we can compare the analysis using the aggregate measures of percent free school meals and average KS1 reading score, with that which adds the individual level variables of free school meals, gender and age (analyses M, N in Appendix A). There is a considerable improvement in prediction when the individual level predictors are added and we shall look at this later in terms of individual schools. The greatest improvement in prediction, however, occurs when the individual pupil level KS1 composite (see Table 1) score is included. In addition, the school average composite score adds to the prediction and when this is done the contribution of the average 1995 school reading score becomes negligible. The average KS1 composite score is dominated by the KS1 reading score (Table 1) and so may be regarded as an approximation to the 1994 school mean for reading, although it does not include those children present in the school at KS1 and not at KS2.

For Mathematics we obtain similar results as for English. The KS1 mean school reading scores are better predictors than the Mathematics scores. The same general conclusions emerge concerning the importance of using individual level prior achievement scores at KS1. Neither writing nor Mathematics for any cohort add to the prediction and the analyses including them are not presented. Nevertheless, in analysis T the composite score coefficient is highly significant. This score has an equal contribution from KS1 reading and Mathematics (Table 1) and can be regarded as approximating the average of these KS1 scores at school level, bearing in mind the same caveats as for English. It is also worth noting in analysis S the random coefficient for free school meals which indicates that the effect of eligibility for free school meals varies across schools. It is also worth noting that while girls make more progress than boys for English, the reverse occurs for Mathematics.

For both English and Mathematics outcome scores, several random coefficient models were fitted, but the only significant model was that including a random coefficient for the composite KS1 score. In other words we see that there is 'differential effectiveness' with schools differing in terms of the progress made by pupils with differing intake (KS1) achievements. We shall look in more detail at this below. There was some evidence for complex (heterogeneous) variation at level 1 with somewhat more variation for boys, but we have omitted this from our analyses for simplicity.

These results emphasise the inadequacy of analyses which fail to use individual level prior achievement data as a predictor of subsequent achievement. If only cross sectional data are used, whether with aggregate or individual level data, it is not possible to make inferences about the 'effectiveness' of schools. Even with longitudinal data of the kind available in this study, care is needed in making inference about school differences, since the use of measures taken at a single prior occasion may not be adequate (Goldstein and Sammons, 1997) and there may be other factors which are important. For example, using a larger data set from Hampshire, Yang et al (1999) showed that SEN status was also an important predictor. There is also the difficult issue of accounting for those students who do not remain in the same school, whose progress may be different from the remainder: this is an issue that affects all predictions, whether using longitudinal data or not (Goldstein, 1997). Finally, the addition of a random coefficient for the KS1 composite score (and additionally FSM for Mathematics) provides an important improvement in prediction since it allows for 'differential effectiveness' where schools can be judged in terms of their 'effects' for students with different initial achievements. As Yang et al show, this provides the basis for providing useful value added feedback to schools at a level of detail which is unavailable otherwise. In the next section we explore some of the differences between the various models in terms of the rank ordering of school effects.

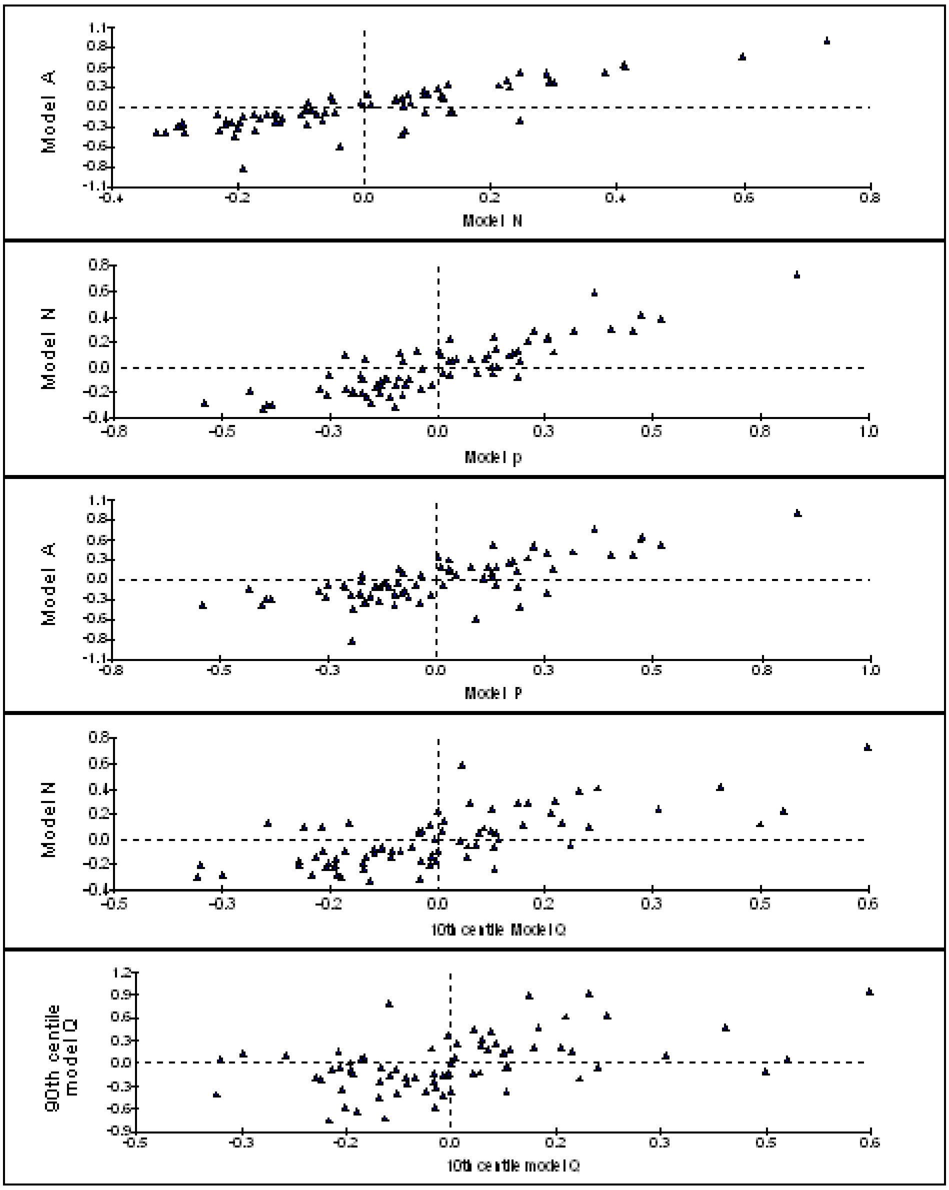
4. School comparisons

We present here comparisons involving completely unadjusted school means (analysis A) with those using cross sectional 1998 data, both individual and aggregate, the model using KS1 individual composite score and the analysis which also includes a random coefficient for the composite score. We first present scatterplots of the 'residuals' for each school, which show how rank orderings change with the model used. We then present some selected plots which identify extreme schools, together with their 90% confidence intervals. The first set of plots are for English.

4.1 English Achievement

All the following models are referred to by the letters assigned to them in Appendix A. Model A contains the unadjusted scores, model N includes only the 1998 variables, both at school and individual level, model P adds the KS1 scores and model Q allows the KS1 individual scores to have a random slope term, the 'differential effectiveness' value added model. As can be seen clearly, compared to the unadjusted scores, even when adjustment is made only for the 1998 variables, some schools are reclassified as less extreme – there is in fact a correlation of 0.85 between residuals from models A and N. Model R, which approximates the use of the 1994 average KS1 reading score, has school residuals which correlate very highly (0.96) with the residuals from model N, and we do not show separate relationships for model R. There is a correlation between the school residuals for model R and those for a model with the student level composite variable instead of the school level variable, of 0.54. This rises to 0.74 when the school level composite variable is added to the model containing just the student level composite variable (model P).

Figure 1. English response models: Residual plots.

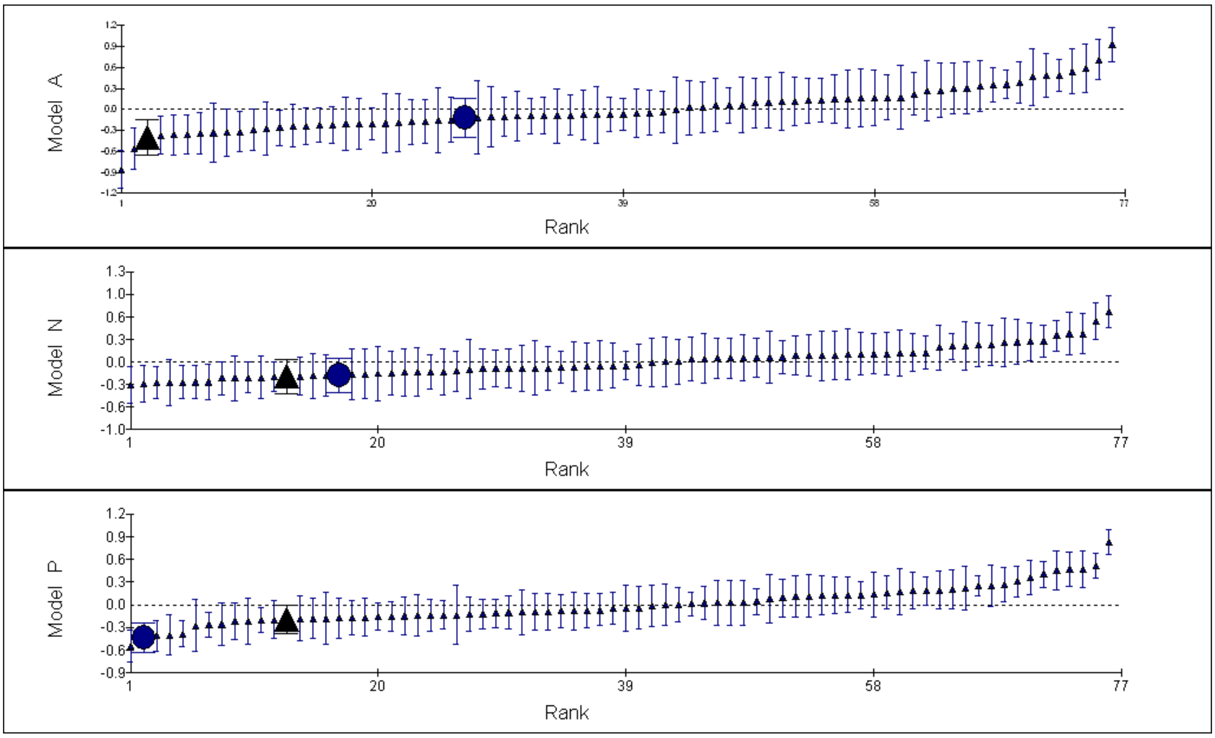


The addition of the KS1 scores (model P) further reclassifies schools from model N which contains just 1998 variables (r=0.87). The full value added analysis which allows for variable ‘slopes’ shows that when low achieving KS1 children (defined as those at the 10th centile of the KS1 composite distribution) are studied, there are larger classification differences compared to the model using just 1998 variables (r=0.70), and a similar result occurs for the high achievers. Finally there is only a relatively weak relationship between the value added scores for the low and high achievers (r=0.48).

One of the reasons for the relatively high correlation between the unadjusted scores and the simple value added scores (model P, r=0.75) is that there is only a moderate difference in the KS1 distributions between schools with an intra-school correlation of 0.15. In this situation it can be shown that a relatively high correlation will tend to occur. This underlines the importance of correctly specifying the value added mode and in this case, as we see, the random slopes model produces very different results. Relying upon only the simple value added model, while better than ignoring the KS1 scores, will allow serious distortions and misleading comparisons.

We now look at some detailed comparisons between schools. For three of our models (unadjusted, using 'proxy' measures, and value added) we have ranked the scores and at the same time provided standard 90% confidence intervals for each score. Where the interval crosses the 'zero' line in Figure 2 we can judge that the school score is significantly different from the LEA mean at the 10% level. Thus, for example, for the unadjusted scores (Model A) 24% satisfy this criterion. If we were to use a 5% level this would reduce to 20%.

Figure 2. English response models: Residual plots with 90% confidence intervals.

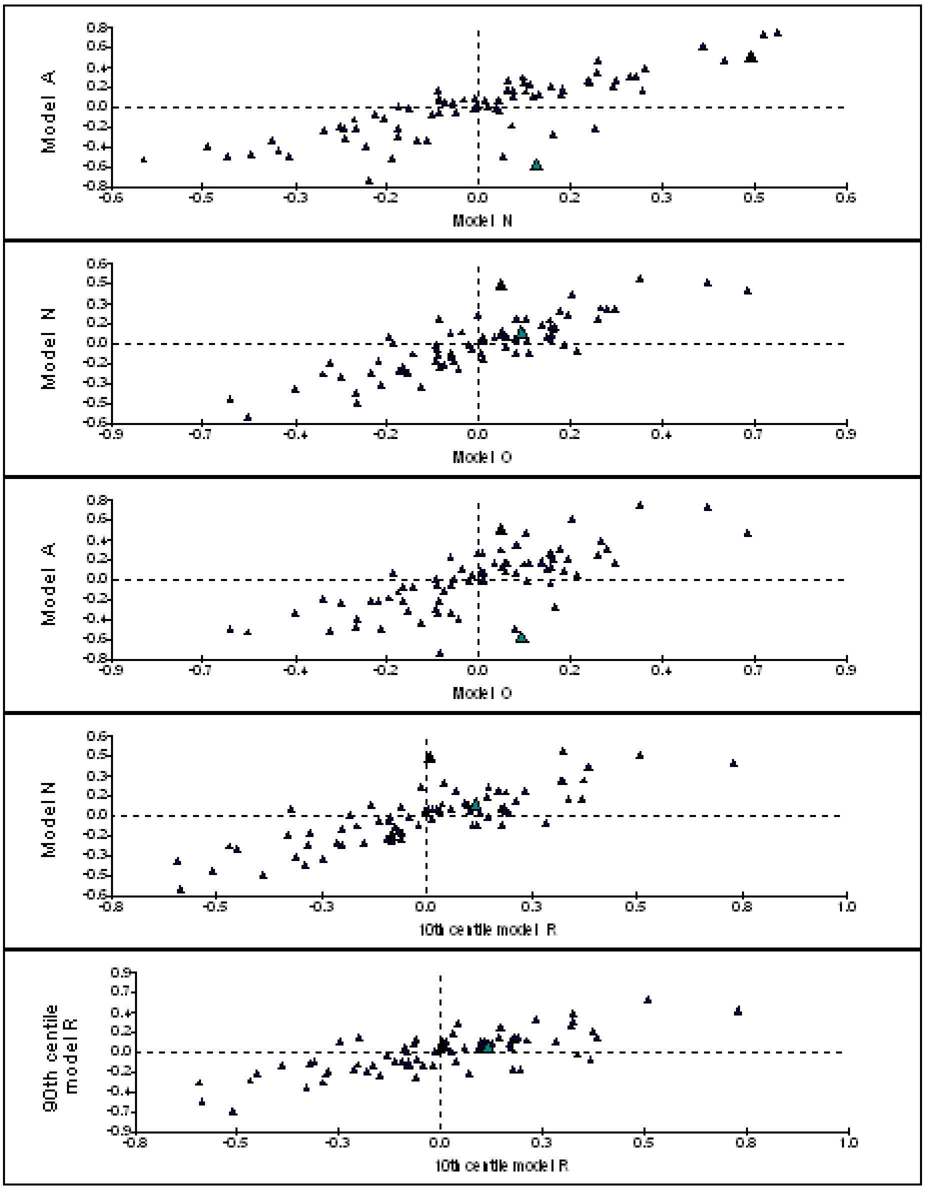


To illustrate one effect of moving towards a more realistic model we have highlighted two particular schools: the one identified by the large triangle appears to be extreme for the unadjusted model and also significantly different from the average, but not so in either of the other models. The one identified by the large circle is not identified as extreme in the unadjusted model, or the model using 1998 variables, but is when a value added analysis is done.

Of those identified as significantly different from the LEA average in model N, about 40% are judged to be not significantly different by model P. Although we do not illustrate them, further examples of such shifts occur when we use the differential effectiveness model and study high and low achievers.

4.2 Mathematics achievement

Figure 3. Mathematics response models: Residual plots.

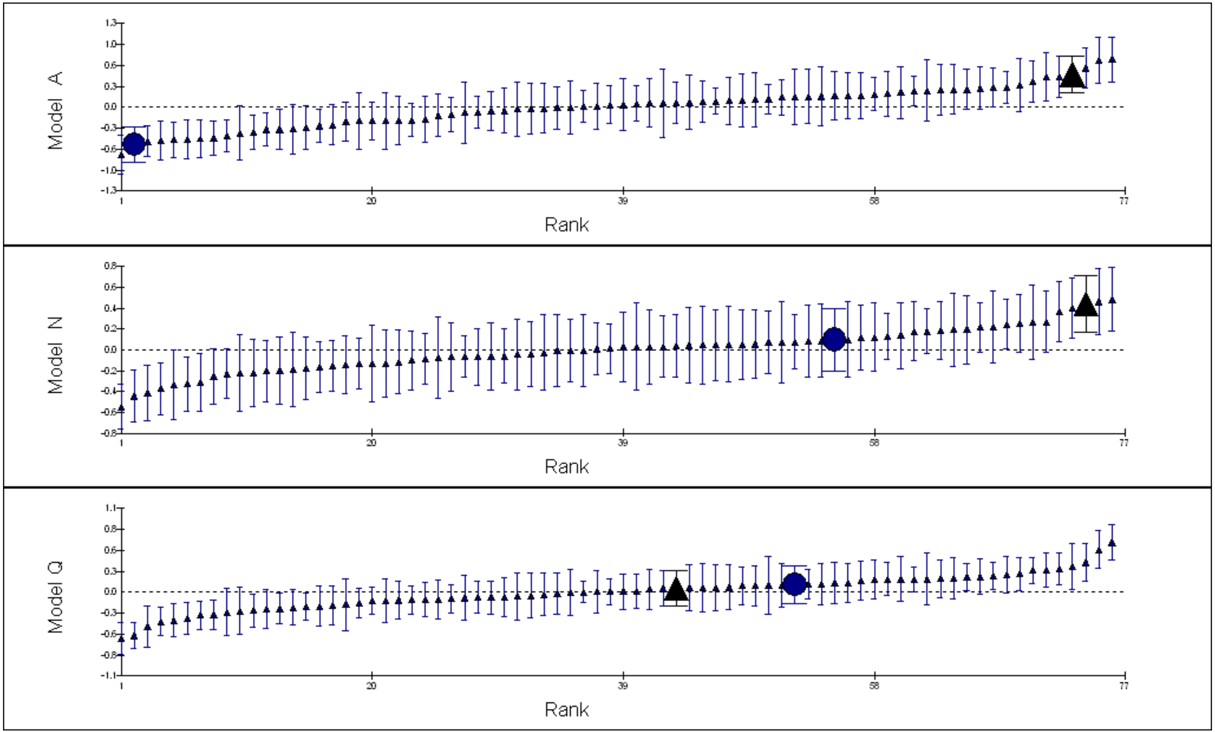


All the models are referred to by the letters assigned to them in Appendix B. As in the case of English we find increasing misclassification as we move from an unadjusted model to the differential effectiveness model. The correlations are similar to those for English, although that between the low and high achievers is somewhat higher (r=0.78).

For model T we have a correlation of 0.85 for the school residuals with those for model N, which is lower than the equivalent correlation for the English response. There is, however, a correlation between residuals from model T and a model with the student level composite score instead of the school level score, of 0.95. This rises to 0.99 when the school level composite score is added to the model with just the student level score (model Q).

These correlations are higher than for English and appear to suggest that, for Mathematics, student level KS1 data may not be necessary for a simple value added analysis which seeks to rank the schools. Nevertheless, the analysis using only school level data (model T) has residuals with lower precision; the confidence intervals are on average 10% wider. Furthermore, it should be remembered that there are important random slopes (models R & S) which do require student level data.

Figure 4. Mathematics response models: Residual plots with 90% confidence intervals.



As with English, we have highlighted examples of schools which are differentially classified by the different models.

5. Conclusions

The analyses we have carried out support, in general terms, what is already known about the use of 'adjustments' which do not involve prior achievement and in particular do not include differential effects (random slopes) (Yang et al., 1999). In particular, we have shown clearly that using readily available 'contemporaneous' data collected at the same time as the KS2 data, whether free school meal eligibility or aggregate KS1 scores, provides an inadequate adjustment. The inclusion of aggregate KS1 scores for earlier cohorts does improve predictions somewhat, but the effect is marginal and these are not adequate substitutions for individual level KS1 scores. It is also worth pointing out that some authors have, confusingly, referred to the use of measures such as free school meals (DfEE, 1997), as providing value added adjustments, whereas of course they cannot do this.

We have also confirmed that even where value added adjustments are made, there still remains considerable uncertainty, as measured by confidence intervals, surrounding estimates for individual schools. This implies that only the extreme schools can be statistically separated from the average. Any comparisons, therefore need to be treated with care. We have argued elsewhere that such comparisons are best regarded as just one indicator of educational quality alongside others (Yang et al., 1999).

In addition, as Goldstein and Sammons (1997) found, adjusting for achievement at a single prior time point is likely to be insufficient, so that in the present case it may well be that including baseline (reception) measures would further refine the value added estimates.

In general the issue of 'model misspecification' is an important one. It is quite possible that even the best available models lack key 'contextualising' variables thus emphasising the additional caution that needs to be taken with interpretations. A particular problem which arises in practice is one that has not been tackled in this research, namely that of pupil mobility. In some schools there is a very high turnover between KS1 and KS2, and in our analyses we have ignored children who do not remain in the same school. Yet we know (Yang et al., 1999) that mobile children tend to make less progress so that to include only those who have been in the school for the whole period will generally lead to biases. In principle, if data were available, the contribution to a child's progress of each school attended, should be incorporated into the models.

6. Policy implications

Regarding the use of 'proxy' measures to inform inspections, our analyses indicate that there is little to be gained by this. They offer only a very small improvement over unadjusted scores and when compared to a full value added analysis provide quite different classifications for many schools. As with the completely unadjusted scores they reflect intake achievements as well as the 'quality' of education within a school. If such proxy measures are made available then, as with the unadjusted key stage results, their limitations need to be set out clearly. This has important implications for the training of inspection teams to understand the issues involved. The results of the present research provide a basis for doing this and also for making suitable modifications to current inspection guidelines so that the limitations of using available data are clearly explained.

While the main focus of our analyses has been on the use of performance indicators in school inspections, our conclusions do of course apply to the general use and interpretation of such data, particularly when they are supplied as published 'league tables.

In the future it is possible that linked individual data will become available on all children so making it possible, in principle, to supply value added analyses for each school and OFSTED has indicated that it wishes to move towards such a scheme. Nevertheless, as the experience in Hampshire has shown, a robust system will be difficult to achieve. Ensuring that all relevant variables are measured and pupil mobility is properly taken into account raises considerable problems. In addition, the inherent uncertainty arising from the relatively small number of pupils involved means that comparisons tend to be useful only for those schools at the extremes. In this context, the value added scheme in Hampshire LEA does provide a viable model where value added data can be used constructively. Extending and developing this scheme on a wider basis is therefore to be welcomed, so long as it serves the formative purposes of school improvement rather than the judgmental one of public league tables.

7. Some caveats

Finally, we need to enter some caution about the present research. The data come from one, partly rural, LEA and within that LEA only from Primary schools. The socio-economic composition of the Hampshire population is not representative of the country as a whole, for example the proportion of pupils eligible for free school meals in our sample is only about 8%

compared to the average for England in 1998 of 20% (DfEE, 1998). Such differences may mean that similar analyses carried out in different areas would show somewhat different patterns.

Nevertheless, analyses of Inner London data of 11 year olds with an average percentage of free school meal eligibility of about 40%, confirm the results of the present research, showing small effects of free school meals when prior attainment is included in the model (Goldstein and Sammons, 1997). Furthermore, the broad findings of our research are in line with other studies (see for example Mortimore et al., 1988) and there is no reason to think that future studies would differ in anything other than some details. It is our view that the conclusions from our research concerning comparisons between unadjusted or proxy data and value added measures, as well as the sizes of associated confidence intervals, are generalisable.

References

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

DfEE (1998). *The Autumn package.* London, Department for Education and Employment.

Goldstein, H. and Fogelman, K. (1974). Age standardisation & seasonal effects in mental testing. *British Journal of Mathematical and Statistical Psychology* **44**: 109-115.

Goldstein, H. and Sammons, P. (1997). The influence of secondary and junior schools on sixteen year examination performance: a cross-classified multilevel analysis. *School effectiveness and school improvement.* **8**: 219-230.

Goldstein, H. (1997). value added tables: the less-than-holy grail. *Managing Schools Today* **6**: 18-19.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., et al. (1988). *School Matters.* Wells, Open Books:

Rasbash, J., Browne, W., Goldstein, H., Yang, M., et al. (1999). *A user's guide to MlwiN (Second Edition).* London, Institute of Education:

Yang, M., Goldstein, H., Rath, T. and Hill, N. (1999). *The use of assessment data for school improvement purposes*. (Oxford Review of Education, to appear).

Appendix A: English KS2 as response

Table 2. Response is KS2 Normalised English test score. (S.E. in brackets).				
Fixed	**A**	**B**	**C**	**D**
Intercept	0.005	0.29	0.25	-0.18
Girl			0.43 (0.05)	0.43 (0.05)
% girls		0.005 (0.004)	0.001 (0.003)	
FSM			-0.38 (0.10)	-0.43 (0.10)
%FSM		-0.032 (0.007)	-0.029 (0.007)	
Random				
Level 2:				
Variance	0.13 (0.03)	0.084 (0.02)	0.084 (0.02)	0.119 (0.03)
Level 1:				
Variance	0.85 (0.03)	0.85 (0.03)	0.80 (0.03)	0.80 (0.03)
-2*log-likelihood	3846.0	3824.9	3742.3	3759.3

Table 3. Response is KS2 Normalised English test score. (S.E. in brackets).				
Fixed	**E**	**F**	**G**	**H**
Intercept	-0.75	-0.63	-0.18	-0.40
Girl	0.43 (0.05)	0.43 (0.05)	0.43 (0.05)	0.43 (0.05)
FSM	-0.38 (0.10)	-0.38 (0.10)	-0.38 (0.10)	-0.39 (0.10)
%FSM	-0.022 (0.007)	-0.023 (0.007)	-0.025 (0.007)	-0.024 (0.007)
Mean KS1 reading 1998	0.33 (0.29)	0.39 (0.21)		
Mean KS1 writing 1998	0.13 (0.37)			
Mean KS1 Mathematics 1998	-0.02 (0.27)			
Mean KS1 reading 1997			0.20 (0.21)	
Mean KS1 reading 1996				0.28 (0.22)
Random				
Level 2:				
Variance	0.08 (0.02)	0.08 (0.02)	0.08 (0.02)	0.08 (0.02)
Level 1:				
Variance	0.80 (0.03)	0.80 (0.03)	0.80 (0.03)	0.80 (0.03)
-2*log-likelihood	3739.1	3739.2	3741.5	3740.7

Table 4. Response is KS2 Normalised English test score. (S.E. in brackets).				
Fixed	**I**	**J**	**K**	**L**
Intercept	-1.01	-1.46	-0.32	-1.12
Girl	0.43 (0.05)	0.43 (0.05)		
FSM	-0.39 (0.10)	-0.39 (0.10)		
%FSM	-0.019 (0.007)	-0.017 (0.008)	-0.025 (0.007)	-0.020 (0.008)
Mean KS1 reading 1998		0.26 (0.22)	0.35 (0.22)	0.23 (0.22)
Mean KS1 reading 1995	0.51 (0.21)	0.44 (0.22)		0.43 (0.22)
Random				
Level 2:				
Variance	0.08 (0.02)	0.08 (0.02)	0.09 (0.02)	0.08 (0.02)
Level 1:				
Variance	0.80 (0.03)	0.80 (0.03)	0.85 (0.03)	0.85 (0.03)
-2*log-likelihood	3736.7	3735.2	3824.8	3821.3

Table 5. Response is KS2 Normalised English test score. (S.E. in brackets).				
Fixed	**M**	**N**		
Intercept	-1.69	-0.73		
Age	0.32 (0.08)	0.32 (0.08)		
Girl	0.43 (0.05)	0.43 (0.05)		
FSM	-0.38 (0.10)	-0.38 (0.10)		
%FSM	-0.016 (0.008)	-0.022 (0.007)		
Mean KS1 reading 1998	0.27 (0.22)	0.39 (0.21)		
Mean KS1 reading 1995	0.43 (0.22)			
Random				
Level 2:				
Variance	0.07 (0.02)	0.08 (0.02)		
Level 1:				
Variance	0.79 (0.03)	0.79 (0.03)		
-2*log-likelihood	3720.3	3742.2		

Table 6. Response is KS2 Normalised English test score. (S.E. in brackets).				
Fixed	**O**	**P**	**Q**	**R**
Intercept	-2.63	-1.70	-1.80	-1.96
Age	-0.13 (0.06)	-0.13 (0.06)	-0.13 (0.06)	0.32 (0.08)
Girl	0.22 (0.03)	0.22 (0.030	0.21 (0.03)	0.42 (0.05)
FSM	-0.22 (0.07)	-0.23 (0.07)	-0.23 (0.07)	-0.37 (0.10)
%FSM	-0.016 (0.007)	-0.019 (0.006)	-0.018 (0.006)	-0.019 (0.006)
KS1 composite score	0.98 (0.03)	1.00 (0.03)	1.01 (0.04)	
School mean KS1 composite		-0.41 (0.12)	-0.40 (0.12)	0.56 (0.12)
Mean KS1 reading 1998	0.08 (0.20)			
Mean KS1 reading 1995	-0.27 (0.21)			
Random				
Level 2:				
Var (intercept)	0.09 (0.02)	0.08 (0.02)	0.31 (0.13)	0.06 (0.02)
Var (composite)			0.04 (0.01)	
Cov (Intcpt., comp.)			-0.10 (0.04)	
Level 1:				
Variance	0.36 (0.01)	0.36 (0.01)	0.35 (0.01)	0.79 (0.03)
-2*log-likelihood	2678.6	2669.3	2642.3	3708.9

Appendix B: Mathematics KS2 as response

Table 7. Response is KS2 Normalised Mathematics test score. (S.E. in brackets).				
Fixed	**A**	**B**	**C**	**D**
Intercept	0.04	0.50 (0.18)	0.49	0.16
Girl			-0.18 (0.05)	0.19 (0.05)
% girls		-0.005 (0.004)	-0.003 (0.004)	
FSM			-0.41 (0.11)	-0.45 (0.11)
%FSM		-0.032 (0.007)	-0.029 (0.007)	
Random				
Level 2:				
Var (intercept)	0.14 (0.03)	0.09 (0.02)	0.09 (0.02)	0.12 (0.03)
Level 1:				
Variance	0.86 (0.03)	0.86 (0.03)	0.84 (0.03)	0.84 (0.03)
-2*log-likelihood	3858.7	3836.4	3806.4	3824.3

Table 8. Response is KS2 Normalised Mathematics test score. (S.E. in brackets).				
Fixed	**E**	**F**	**G**	**H**
Intercept	-0.88	-1.09	-0.17	-0.97
Girl	-0.18 (0.05)	-0.19 (0.05)	-0.19 (0.05)	-0.19 (0.05)
FSM	-0.41 (0.11)	-0.41 (0.11)	-0.41 (0.11)	-0.42 (0.11)
%FSM	-0.021 (0.007)	-0.020 (0.007)	-0.025 (0.007)	-0.021 (0.008)
Mean KS1 reading 1998	0.44 (0.29)	0.32 (0.25)		
Mean KS1 writing 1998	-0.28 (0.36)			
Mean KS1 Mathematics 1998	0.36 (0.27)	0.32 (0.27)		
Mean KS1 reading 1997			0.44 (0.31)	
Mean KS1 Mathematics 1997			-0.22 (0.28)	
Mean KS1 reading 1996				0.47 (0.26)
Mean KS1 Mathematics 1996				0.11 (0.30)
Random				
Level 2:				
Var (intercept)	0.08 (0.02)	0.08 (0.02)	0.08 (0.02)	0.08 (0.02)
Level 1:				
Variance	0.84 (0.030	0.84 (0.03)	0.84 (0.03)	0.84 (0.03)
-2*log-likelihood	3799.9	3800.5	3804.7	3800.8

Table 9. Response is KS2 Normalised Mathematics test score. (S.E. in brackets).				
Fixed	**I**	**J**	**K**	**L**
Intercept	-1.35	-2.05	-1.92	-1.91
Girl	-0.19 (0.05)	-0.19 (0.05)	-0.19 (0.05)	-0.20 (0.05)
FSM	-0.41 (0.11)	-0.41 (0.11)	-0.41 (0.11)	
%FSM	-0.017 (0.007)	-0.012 (0.008)	-0.013 (0.007)	-0.017 (0.007)
Mean KS1 reading 1998		0.19 (0.25)		
Mean KS1 Mathematics 1998		0.24 (0.26)	0.34 (0.22)	0.35 (0.23)
Mean KS1 reading 1995	0.68 (0.25)	0.60 (0.21)	0.63 (0.21)	0.63 (0.21)
Mean KS1 Mathematics 1995	0.03(0.20)			
Random				
Level 2:				
Var (intercept)	0.07 (0.02)	0.07 (0.02)	0.07 (0.02)	0.07 (0.02)
Level 1:				
Variance	0.84 (0.03)	0.84 (0.03)	0.84 (0.03)	0.85 (0.03)
-2*log-likelihood	3795.7	3792.8	3793.4	3808.6

Table 10. Response is KS2 Normalised Mathematics test score. (S.E. in brackets).				
Fixed	**M**	**N**	**O**	**P**
Intercept	-1.10	-1.15	-3.00	-3.99
Age		0.25 (0.08)	-0.16 (0.06)	-0.16 (0.06)
Girl	-0.20 (0.05)	-0.19 (0. 05)	-0.29 (0.04)	-0.29 (0.04)
FSM		-0.41 (0.11)	-0.26 (0.07)	-0.26 (0.07)
%FSM	-0.023 (0.007)	-0.020 (0.007)	-0.012 (0.007)	-0.012 (0.007)
KS1 composite score			1.02 (0.03)	1.03 (0.03)
School mean KS1 composite				-0.43 (0.14)
Mean KS1 reading 1998	0.33 (0.26)	0.32 (0.25)	0.13 (0.23)	0.15 (0.22)
Mean KS1 Mathematics 1998	0.32 (0.27)	0.33 (0.27)	-0.15 (0.24)	0.02 (0.24)
Mean KS1 reading 1995			0.08 (0.20)	0.30 (0.20)
Random				
Level 2:				
Var (intercept)	0.08 (0.02)	0.08 (0.02)	0.08 (0.02)	0.07 (0.02)
Level 1:				
Variance	0.85 (0.03)	0.84 (0.03)	0.40 (0.02)	0.40 (0.02)
-2*log-likelihood	3815.4	3791.5	2802.0	2793.3

Table 11. Response is KS2 Normalised Mathematics test score. (S.E. in brackets).				
Fixed	**Q**	**R**	**S**	**T**
Intercept	-1.86	-1.80	-1.76	-1.92
Age	-0.16 (0.06)	-0.17 (0.06)	-0.17 (0.06)	0.25 (0.08)
Girl	-0.29 (0.04)	-0.30 (0.03)	-0.30 (0.03)	-0.20 (0.05)
FSM	-0.26 (0.04)	-0.26 (0.07)	-0.33 (0.10)	-0.39 (0.11)
%FSM	-0.019 (0.006)	-0.019 (0.006)	-0.018 (0.006)	-0.018 (0.006)
KS1 composite score	1.03 (0.03)	1.05 (0.03)	1.05 (0.03)	
School mean KS1 composite	-0.32 (0.13)	-0.35 (0.13)	-0.37 (0.13)	0.70 (0.13)
Random				
Level 2:				
Var (intercept)	0.07 (0.02)	0.32 (0.14)	0.34 (0.14)	0.05 (0.02)
Var (composite)		0.02 (0.01)	0.02 (0.01)	
Var (FSM)			0.16 (0.08)	
Cov (Intcpt., comp.)		-0.08 (0.04)	-0.08 (0.04)	
Cov (Intcpt., FSM)			-0.07 (0.09)	
Cov (comp., FSM)			0.002 (0.03)	
Level 1:				
Variance	0.40 (0.02)	0.39 (0.02)	0.38 (0.02)	0.84 (0.03)
-2*log-likelihood	2796.7	2788.0	2775.1	3772.9