

Do Specialist Schools Add Value? – Some methodological problems

By

Ian Schagen, NFER (i.schagen@nfer.ac.uk)

Harvey Goldstein, Institute of Education (h.goldstein@ioe.ac.uk)

Introduction

The impact of different school types on the progress made by their pupils is an important topic, with especial relevance to the ongoing debate about the value of the specialist schools programme (see Chitty, 2001; Croxford, 2001). With the advent of detailed and exhaustive national value-added datasets it would seem that it should be a plain and simple matter to tackle these questions using objective information about the progress made by just about every pupil in the country in different types of schools. Unfortunately, like the truth, this is ‘never plain, and rarely simple’, and it is clear from some recent attempts to carry out this kind of analysis that issues of methodology are crucial and need to be debated in a public forum.

In this article we will focus on analysis that claims to demonstrate superior progress in specialist schools as compared with standard comprehensives, as presented by Jesson with Taylor (2002). The issue is not so much about the findings but about the methodology used. Some of the same problems exist in much of the ‘value-added’ analysis which is current, whether intended to address differential progress for different school types, general information for school improvement or via commercial packages used by schools.

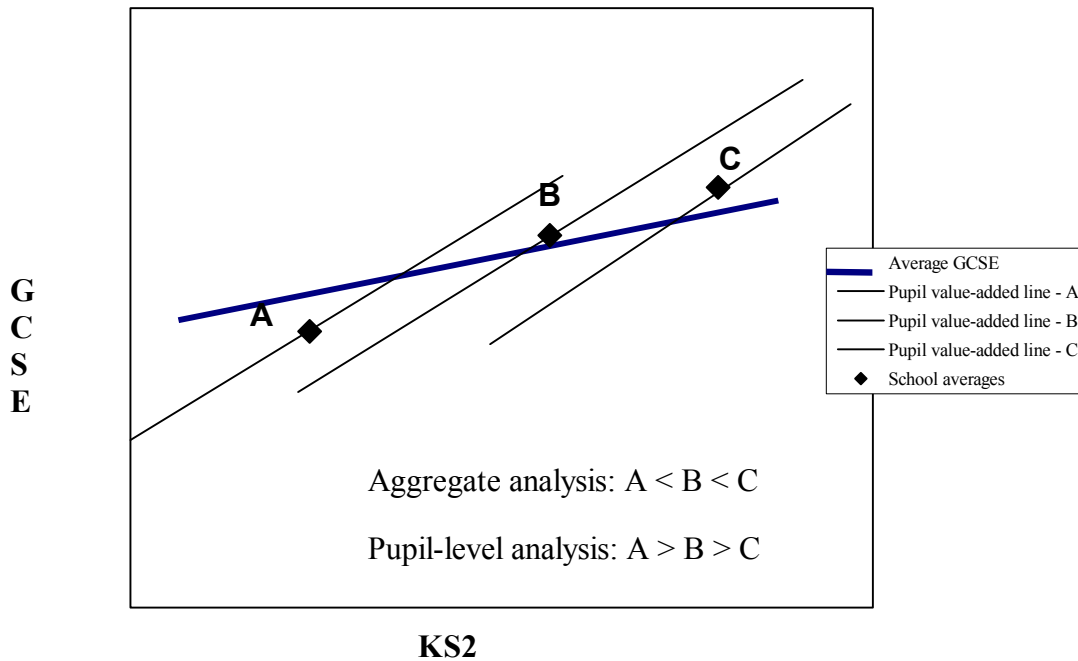
Analysis of national value-added datasets

Jesson (2002) reports analysis of an enhanced national value-added dataset linking KS2 results in 1996 to GCSE outcomes in 2001. It contains information on around 510,000 pupils and is thus an impressive collection of data. However, the quality of the results extracted from a dataset depends on both the quality of the data itself and the validity and appropriateness of the analysis carried out. There is general agreement among academics involved in this kind of work (see Goldstein, 1997; Saunders, 1999) that best practice in value-added analysis is that it should:

- Be based on the analysis of pupil-level data.
- Use multilevel modelling, which is technically efficient and takes proper account of background factors and variations at school and pupil levels.
- Take account of all relevant contextual factors, such as percentage eligible for free school meals.

Sadly, Jesson’s work fails on all these counts. Although he starts with pupil-level data, the analysis seems to have been carried out at the school level (this is shown by the fact that he uses as an outcome the percentage achieving 5+ A* to C GCSE grades – a school-level variable). The following illustration shows how the use of school level analysis can produce misleading results. This diagram shows how this can arise. It uses three individual schools, but applies similarly to different school types such as specialist/non-specialist.

Comparison of Aggregate versus Pupil-level Analysis for 3 Schools



The thick line is the relationship between the *average* GCSE and *average* KS2 scores; the GCSE mean for school A lies below this line so it ‘does less well’ than B which lies above the line, and B ‘does less well’ than C which is even further above the line.

The thin lines are the correct pupil-level value-added lines for each school; for each KS2 score, where there is overlap of KS2 scores across the schools, pupils in school A do better than those in school B who do better than those in school C. In other words, the aggregate analysis gives the opposite result from the pupil-level analysis.

Of course, there are situations where both aggregate and individual level analyses yield similar conclusions, but this can only be established by carrying out both types of analysis on the same data set. In the absence of strong evidence that this is the case, we should exercise great caution in basing any conclusions on an aggregate analysis. In addition, much of the school effectiveness literature implies that the lines representing the individual-level relationships, rather than being parallel (as shown in the diagram), are far from parallel and that such ‘interactions’ are important features of any inferences.

Jesson is not alone in arguing against the use of multilevel modelling, despite the fact that it has transformed the ability of statisticians to carry out analysis of educational data and can handle, potentially, data structures of considerable complexity. His reasons for not using it seem rather muddled, and his own methods of analysis and presentation of results are far from clear. Furthermore, we believe that most researchers would agree that a correct analysis was superior to a simple but incorrect one. In practice a sophisticated and valid analysis can have its results presented clearly and accessibly, with a little care. In fact, it is precisely the ability of multilevel modelling to expose any real intrinsic complexity that enables the user to present results at a level of clarity that does not distort the underlying reality.

One of the big advantages of multilevel modelling is the fact that it not only computes estimates of the effects involved (e.g. the effect of specialist schools) but also the uncertainty in those estimates. This is very important, as it allows us to report 'confidence intervals' for our findings and prevents us from giving results which are simply the result of chance variation. Jesson's paper produces a lot of numbers, graphs and formulae, but nowhere is there any acknowledgement of uncertainty or presentation of confidence intervals.

Jesson's failure to take account of school context, as measured by free school meals, also makes his results less robust. Research has consistently shown that this has an impact, even on value-added results (see e.g. Saunders, 1999; Schagen & Morrison, 1998). Furthermore, he distinguishes between specialist and non-specialist schools, but makes no distinction between different types of specialism, nor does he discuss the date at which they became specialist. In recent work by Schagen *et al.* (2002) significant differences between the apparent impacts of different types of specialist school were detected, and it seems important that this should be recognised.

Interpretation of results

Despite the above critique of the analysis methods used by Jesson in approaching value-added datasets, it is of course possible that the results obtained may not be too far away from those derived from a more correct analysis. For example, Jesson (2002, page 5, Table 4) seems to show a total GCSE point score value-added advantage of about 2.2 points in favour of specialist schools. In analysis of previous datasets using multilevel models and controlling for school context, values ranging from zero to 1.5 for different types of specialism were obtained (Schagen *et al.*, 2002, Table 3.2, p.21). Even so, the question of the interpretation of these results is one that must be faced.

For Jesson the finding of a positive value-added effect associated with specialist schools is proof of the effectiveness of the specialist school programme and of the 'success' of these schools. However, statisticians should be rightly cautious about attributing causality on the basis of such findings. It could be that the specialist nature of these schools has brought about a small but significant improvement in their value-added performance. On the other hand, schools have to bid for specialist status, and we understand that one of the factors taken into account is the school's performance in value-added terms. Furthermore, schools have to provide evidence of sponsorship, and this is probably easier for those with more middle-class and supportive parents and local community – factors which will generally affect value-added performance. So an alternative hypothesis is that schools become specialist because they have a tendency to achieve better value-added results, although other research (Gray *et al.*, 2001) suggests that the correlations between value added results from year to year are only moderate.

A further issue is that of pupil mobility. It may well be the case that pupils making the most progress tend to move from comprehensives into specialist schools, so boosting the value added performance of the latter. None of the datasets analysed so far casts light on such a possibility. The point is not so much that any of these alternative hypotheses are the correct explanation for the 'specialist school effect' detected by Jesson, but that no alternative explanations have been considered.

Conclusions

We have argued that the important educational policy issue of the effectiveness of specialist schools needs to be informed by research of the highest quality. Such research should be based on the best quantitative analysis techniques (i.e. multilevel modelling of pupil-level results) and should take account of all relevant contextual information. The interpretation of such analysis should consider all possible explanations of the findings before delivering a judgement on the effectiveness of a particular initiative. Ideally, such research should be subjected to peer review prior to publication, even where time constraints may make this difficult. An open debate within the educational research community would be the best way of ensuring that results which feature in a government white paper as part of an argument in favour of specialist schools (England. Parliament. HoC, 2001) are well-founded.

(This paper has been submitted to BERA 'Research Intelligence' for publication.)

References

CHITTY, C. (2001) Modernisation spells destruction, *Education Review*, 15(1), pp. 37–44.

CROXFORD, L. (2001) School differences and social segregation: comparison between England, Wales and Scotland, *Education Review*, 15(1), pp. 68–73.

ENGLAND. PARLIAMENT. HOUSE OF COMMONS (2001). *Schools Achieving Success* (Cm.5230). London: The Stationery Office.

GRAY, J., GOLDSTEIN, H. & THOMAS, S. (2001) Predicting the future: the role of past performance in determining trends in institutional effectiveness at A level, *British Educational Research Journal*, 27(4), pp. 391–405.

GOLDSTEIN, H. (1997) Methods in school effectiveness research, *School Effectiveness and School Improvement*, 8(4), pp. 369–395.

JESSON, D. (2002) Progress of pupils in specialist and other schools from key stage 2 to GCSE, In: D. JESSON & C. TAYLOR *Value Added and the Benefits of Specialism* [online]. Available: <http://www.tctrust.org.uk/news/valueadded2002.html> [7 June, 2002].

JESSON, D. with TAYLOR, C. (2001) *Educational Outcomes and Value Added Analysis of Specialist Schools for the Year 2000* (London, Technology Colleges Trust).

SAUNDERS, L. (1999) A brief history of educational 'value added': how did we get to where we are? *School Effectiveness and School Improvement*, 10(2), pp. 233–256.

SCHAGEN, S., DAVIES, D., RUDD, P. & SCHAGEN, I. (2002) *The Impact of Specialist and Faith Schools on Performance* (LGA Research Report 28) (Slough, NFER).

SCHAGEN, I. & MORRISON, J. (1998) QUASE: Quantitative Analysis for Self-Evaluation. Overview Report 1997: Analysis of GCSE Cohorts 1994 to 1996 (Slough, NFER).