(Oxford review of Education, 1998, **24**, 521-523)

A Response to Gibson and Asthana

Gibson and Asthana (Oxford Review of Education, 1998, 24, 195-210)present a critique of school effectiveness research (SER) based upon the following two propositions:

- That the focus of SER and its specific conclusions have reinforced government policies concerned with identifying schools as the sole agents of 'success' and 'failure'
- That SER has ignored the strong relationship between social background and school performance

The first proposition is concerned with the politics of SER whereas the second is more of a technical issue. Both are contentious and I would like to examine each in turn.

First, I have a great deal of sympathy with the authors' first point. There can be little doubt that the findings and terminology of SER have been 'hijacked', especially by the New labour administration, for the kinds of political purposes to which Gibson and Asthana object. This is nicely exemplified by one of their quotations from the 1997 White paper

'schools with similar intakes of pupils achieve widely differing results'.

In fact this statement is based upon an analysis presented in the Literacy Task Force report (Barber, 1997) which was an aggregate level analysis relating the average Key Stage 2 test scores for a sample of schools to the percentage of pupils having free school meals in those schools. The Task Force report and the White paper show a misunderstanding of SER, both because such an aggregate level analysis tells us little about effects on *students* (more about this below) and because adjusting for free school meal uptake is *not* the same as adjusting for intake - by which the SER literature means, at the very least, prior achievement.

Yet it is one thing to complain about such political abuse of research findings and quite another to criticise the research field for allowing this to happen. Of course, it may well be appropriate to castigate particular individual SE researchers who have moved directly into government or played a role in its agencies, but this does not amount to a balanced critique of the whole research area.

There are many problems with SER, some of which I have discussed elsewhere (Goldstein, 1997). The statistical models used and the data available are often crude and there are severe constraints upon the strengths of the inferences which may legitimately be drawn from analyses. As the research progresses people are beginning to discover that much of the interesting variation occurs at class and teacher level, as Bishop and Asthana point out, and also that schools are differentially effective for different kinds of students (Nuttall et al., 1989). It is also fairly clear that we need to take account of more than a simple measure of prior achievement to carry out a proper contextualisation or adjustment (Goldstein and Sammons, 1997). The trouble with the Bishop and Asthana paper is that it misses the mark by concentrating on the political abuse of SER and ignoring other insights about schooling that are beginning to emerge.

As for Bishop and Asthana's second proposition, it is quite clear that there is a strong association between social background and school performance. It is, however, more complex than these

authors describe. While it is true that most of the variation between school averages can be explained by the percentage of free school meal eligibility or other measures of social background, the same is the case for prior achievement measures. For example, using the same data as analysed by Goldstein et al. (1993) one obtains an *aggregate level* correlation between a combination of prior achievement measures (reading scores and verbal reasoning groups) and GCSE scores of 0.95 - rather higher than those quoted for social background by Bishop and Asthana. This is, of course, simply another example of the ecological fallacy as described by the authors themselves. Such correlations are not relevant to questions about the relative influences of social and other factors on pupil achievement. What we need to know is something about the contribution of social background after adjusting for prior achievement, and vice-versa on the basis of a *student level* analysis.

Using this same data set we can look at free school meals eligibility, measured on individual pupils and also in terms of the percentage in the school. What we find is that, based upon a multilevel model which simultaneously incorporates school level and pupil level effects, the free school meals variables account for about 20% of the between-school variation. If we now add in the achievement scores we can account for a total of 55% of the between-school variation; note that this is a more modest figure than before because we are properly modelling all sources of variation at the same time as they relate to student achievements, not merely relating school means. If now we reverse the procedure and just fit the prior achievement measures on their own we still account for almost 55% of the between-school variation! There remains an additional overall effect of eligibility for free school meals, but this variable does not account for any more of the differences *between* schools - only for that between pupils *within* schools. Other published SER has looked at the additional effect of social background type measures with similar results. It is simply not true to say, therefore, that SER has neglected such factors. Of course, prior achievement itself is related to social background, but as the figures I have quoted suggest, it is not necessarily the most important factor. The point of SER is to try to establish what factors are associated with school differences *after adjusting for the factors present at the start of the relevant phase of schooling*.

Finally, Bishop and Asthana seem to fall into the same trap as the government in their interpretation of SER as being concerned with the ranking of individual schools. The use of SER to identify individually 'effective' or 'failing' schools is really a side issue for research. To begin with, the sampling variation associated with typical value added scores is so large that most schools, let alone departments, cannot be statistically separated. At best, the ranking of schools is a screening device to be used for school improvement purposes by schools and Leas and is not a reliable accountability device to be published as if the rankings represented any kind of accurate statement about the quality of education. Unfortunately, too much SER (my own included) has been sidetracked down this path. The real value of SER is in furthering our understanding of the complexities of educational structures and how they come to influence pupils and teachers. It would be a pity if such work was ignored either because of its abuse by government or because of the kinds of misunderstandings presented in the Bishop and Asthana paper.

Harvey Goldstein

18 December 1998

References

Barber, M. (1997). *A reading revolution: how we can teach every child to read well*. London, Institute of Education (Produced for the Labour Party).

Goldstein, H. (1997). Methods in school effectiveness research. *School effectiveness and school improvement.* **8**: 369-95.

Nuttall, D. L., Goldstein, H., Prosser, R. and Rasbash, J. (1989). Differential school effectiveness. *International Journal of Educational Research* **13**: 769-776.

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., Rasbash, J., Yang, M., Woodhouse, G., et al. (1993). A multilevel analysis of school examination results. *Oxford Review of Education* **19**: 425-433.