

## **The 2001 Education White Paper and evidence based policy: a commentary. (October 2001).**

By

Harvey Goldstein

Institute of Education, London

[h.goldstein@ioe.ac.uk](mailto:h.goldstein@ioe.ac.uk)

### **Introduction**

Laudably, the present government has committed itself to using evidence as a basis for policy, not least by funding a unit devoted to reviewing relevant research evidence, located at the Institute of Education. The 2001 White paper ([www.dfes.gov.uk/achievinguccess/](http://www.dfes.gov.uk/achievinguccess/)) provides one opportunity to judge whether this stated commitment is matched by reality.

I will look briefly at what is said about school performance tables, where there is now a considerable body of evidence about their validity and follow with an analysis of the major new initiative in the White Paper, the introduction of specialist schools.

### **Performance tables**

The White paper discusses the introduction of 'value added' rankings of schools in performance tables (para 3.4), first for KS3 to GCSE and later for KS2 to GCSE. The evidence from school effectiveness research (for a review download the article 'League tables and Schooling' at [www.ioe.ac.uk/hgpersonal/](http://www.ioe.ac.uk/hgpersonal/)) is that there are three major issues that need to be confronted. First, it is important to take account of achievement prior to KS2 (in the case of KS2 to GCSE comparisons among schools), since, for example, junior school attended and achievement at the start of junior school have an effect on GCSE over and above achievement at KS2.

Secondly, there is much pupil mobility between KS2 and GCSE so that many, and in some cases most, pupils taking GCSE in a school will not have started their secondary schooling in the same school. This means that their 'value added' score, where it can be computed, cannot be ascribed fully to their GCSE school. The solution to this problem lies in an analysis that correctly apportions the score among the schools attended, but this is quite complicated and requires very good data on transitions.

The third problem is that even were valid value added scores can be computed for a set of schools, the uncertainty associated with them, because of the relatively small numbers of pupils involved, is large. This implies that most schools cannot reliably be separated, rendering rankings rather meaningless. This does not mean that the identification of 'outlying' schools is pointless (for a discussion download the article 'The use of value added information in judging school performance' at [www.ioe.ac.uk/hgpersonal/](http://www.ioe.ac.uk/hgpersonal/)) but it does imply that performance tables cannot be used for routine comparisons or to inform parental choice in a reliable fashion.

The white paper in fact discusses none of these issues adequately, and this is disturbing since they have been in the public domain for some time, have been the subject of a report to OFSTED (see previous reference) and a seminar to members of Parliament (Parliamentary and Scientific Cttee., January 2001).

## Specialist schools

Chapter 5 is devoted to justifying the introduction of specialist schools; there are planned to be 50% of secondary schools in this category by 2005, and these schools will be selected in terms of their existing 'performance' (it is not clear how this will be judged). One concern is that these schools may undermine the comprehensive principle by covertly acquiring the status of selective schools. The white paper response does not even attempt to present evidence merely remarking "There are those who have said that specialist schools will create a two tier system. They won't."

The chapter does, however, claim to present evidence for the superiority of specialist schools in terms of achievement, using a study which compares existing specialist schools with the remainder. In fact the White Paper actually justifies this policy with the claim that "specialist schools are a key part of our proposals for a more diverse system because of their *proven success in raising standards*" (para. 5.9) (my italics).

The evidence arises from a study carried out for the Technology Colleges Trust (download from [www.tctrust.org.uk](http://www.tctrust.org.uk)) and published in Summer 2001. It carries out two sets of 'value added' analyses, one charts progress from KS3 to GCSE and one from KS2 to GCSE. In fact, neither analysis was a true value added one since they were carried out not using pupil data but taking the average KS3 or KS2 scores for each school and comparing these with the average GCSE scores. Even though, for the KS3 analysis, the same pupils are involved, such aggregate level 'pseudo' value added analyses may be misleading for purposes of comparing types of schools (see Woodhouse and Goldstein, 1989 for a discussion of this). To be fair to the authors of the study, they do state in a footnote (P30) that the aggregate analysis 'potentially obscures information about how each school deals with the variety of pupils it educates'. Nevertheless, the report does *not* stress that such aggregate analysis can give incorrect results. The hypothetical diagram below illustrates how this can arise. It uses three individual schools, but applies similarly to different school types such as specialist/non-specialist.

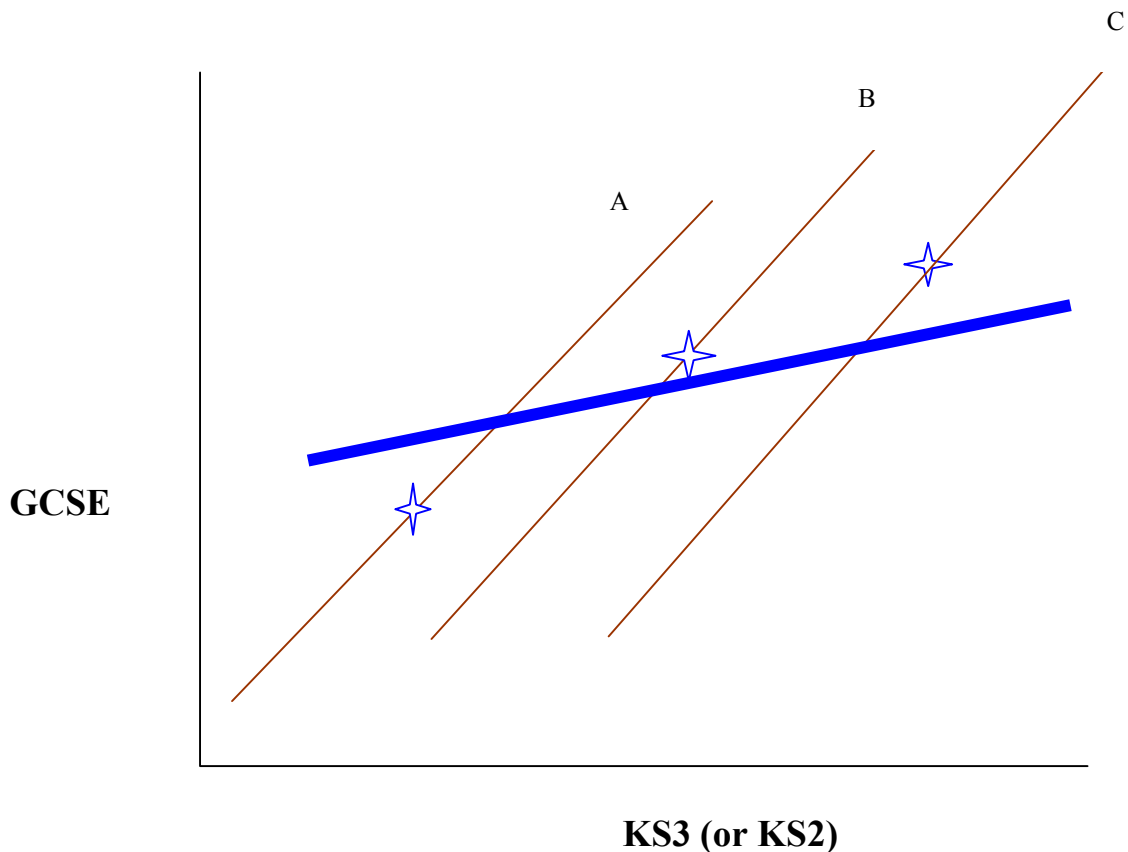
Comparison of GCSE for *given* KS3 (or KS2) scores

Aggregate vs pupil level value added for 3 schools.

Aggregate analysis:  $A < B < C$

Pupil level analysis:  $A > B > C$

★ Indicates school average



The thick (blue) line is the relationship between the *average* GCSE and *average* KS3 (or KS2) scores; the GCSE mean for A lies below this line so 'does less well' than B which lies above and B does less well than C which lies further above.

The thin (brown) lines are the correct pupil level 'value added' lines; for each KS3 (or KS2) score, pupils in A do better than those in B who do better than C.

In other words the aggregate level 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 shows that the lines representing the individual level relationships are far from parallel (as shown in the

diagram) and that such 'interactions' are important features of any inferences. The other problem with analysing progress between KS3 and GCSE is that this represents only part of the secondary school period and ignores changes between KS2 and KS3.

The report also carries out an analysis of KS2 to GCSE results and shows, as with the KS3 to GCSE analysis, that the specialist schools do better than others. The report claims that 'irrespective of intake differences specialist schools had 53% A\*-C passes at GCSE compared to 43% for remainder'.

A further problem arises with the KS2 analysis in that the average KS2 score for a school is based upon a different set of pupils to the average GCSE score because of mobility, as explained above. In particular, it may well be the case that the specialist schools in the study acquired high achieving pupils from nearby comprehensives during the period under study so boosting their pseudo value added scores. For both these reasons the results of the analysis are unsafe and should not be used to draw conclusions about the relative merits of different kinds of schools. Thus the study report's introduction would seem to be misleading in claiming that, because specialist and non-specialist schools have similar KS2 average scores, the better average performance of the former at GCSE is not dependent on selective intake differences.

Similarly, there is an analysis of changes in the GCSE performance of schools from 1999 to 2000, without taking into account possible changes in intakes and this also is an insufficient basis for drawing conclusions. In short, the 'evidence' for the 'proven success' of specialist schools does not stand up to close examination.

This study of specialist schools was not published in a peer-reviewed journal nor, it seems, subject to peer review via seminars or conferences. This is not necessarily a criticism of the authors of the report, since many research results become public through reports to sponsors. It does, however, raise an important issue for those who would wish to use the findings, which may be controversial and open to technical criticism. The role of peer review is to air such issues and it is reasonable to assume that had the report been so exposed its weaknesses would almost certainly have been discussed, and even perhaps accepted by the authors, and there would have been doubts expressed about drawing any firm conclusions. It is not clear whether the authors of the White Paper sought views on the adequacy of the research before using it, but again there are those within the DFES itself who would have cautioned against taking the results of the study at face value. Given that the research supported what was already Government policy, it would seem that this is what drove the decision to use it as 'evidence'.

If government is serious about the use of evidence in policy formation it surely needs to avoid such opportunistic use of research. At the very least it should expect to use research that has been through a validation process, typically by stringent peer review and where possible by means of replication. Indeed, it could be argued that government should *require* some kind of review process before any results are regarded as 'evidence'. Needless to say, this present case does not seem to augur very well for the future. If they wish to engage people in serious rational debates, those responsible for formulating policy will need to exercise considerably more respect for soundly based evidence than is so far apparent.

## Reference

Woodhouse, G. and Goldstein, H. (1989). Educational Performance Indicators and LEA league tables. *Oxford Review of Education* 14: 301-319.