

(To appear in *Parliamentary Brief*; Winter 1998)

Targets, performance indicators and the proper uses of educational data.

By

Harvey Goldstein

Institute of Education

University of London

h.goldstein@ioe.ac.uk

www.ioe.ac.uk/hgoldstn/

League tables

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. Since 1996 the Conservative government, followed by New Labour has extended such 'league tables' to key stages 1, 2 and 3, at both local and national level. These tables contain average, unadjusted, test and exam scores for each school. At the same time the Government has indicated that it is moving to the publication of so called 'value added' or 'intake adjusted' scores and a project under the auspices of the Qualifications and Curriculum Authority (SCAA, 1997) has studied the implementation of such a scheme nationally. Also, the government has given each LEA performance 'targets' and individual schools also have targets to meet; the targets being defined in terms of the percentages of pupils gaining particular levels at each key stage. The targets at Key Stages 2 and 4 are public knowledge, at least at local level, and hence perform the same function as the unadjusted performance tables in that they provide a crude ranking of schools as measured by the extent to whether the targets are met.

The principal argument against these unadjusted 'league tables' and likewise the targets is that the performance of a school is determined largely by the pre-existing achievements of those pupils 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. Hence, *at the very minimum* there has to be an adjustment based upon the intake achievements of pupils. This adjustment, leading to 'value added' scores for schools, is now a well established principle and accepted formally by the government. If schools are compared on a 'value added' basis then quite different rankings are obtained compared with a simple unadjusted comparison. In the remainder of this article I will discuss how far it is possible to use such value added scores for making such comparisons. I will point out the severe limitations that must be placed upon any *published* rankings and will argue that they are only of any real use if used directly by schools for improvement purposes without public exposure

in the form of rankings. A full technical discussion of these issues can be found in Goldstein and Spiegelhalter (1996). First, let me say something about the confusions inherent in current government policies.

In the latest DfEE 'Autumn Package' circulated to all schools in November 1998, schools are advised to compare themselves in terms of average key stage test scores with unadjusted national norms. They are also advised to carry out simple 'value added' comparisons. Yet there is no discussion of the difference between these and where different results are obtained the schools are given no advice as to how to interpret these. The same ambiguity is apparent in the government's adherence to the publication of crude unadjusted league tables, without caveats or health warnings, while officially admitting implicitly that these are invalid in comparison with value added analyses. In fact the situation is even worse, since the government is well aware that all rankings of schools using key stage results are very imprecise and most schools simply cannot be separated in a scientifically acceptable way - that is the league tables are misleading when they provide such rankings. It is not my intention in this article to explore reasons for such apparently deliberate attempts to ignore research evidence and to provide misleading information, but it is worth pointing out that the previous and the present administration have failed to provide any logically coherent defence for their positions on this matter. It is also worth remarking that in other systems, for example New South Wales in Australia, the deficiencies of public rankings are accepted with the result that it is now illegal for anyone to publish such league tables.

Value added comparisons

Several LEAs have been working for some time now on value added systems which allow schools to make detailed comparisons of their pupils' achievements in different subject areas. A feature of all these schemes is that the results are fed back directly to schools, together with all the necessary caveats so that they can be used *in conjunction with other locally sensitive and properly contextualised data* as an aid to school improvement. The results, where possible, also report comparisons separately for different kinds of pupils, such as low achievers or boys separately from girls. Hampshire LEA and the London Institute of Education have been collaborating on such a scheme of value added in Primary schools. The study has looked at individual pupil achievement between baseline and KS1 and between KS1 and KS2. In addition a large number of school and pupil level variables have been analysed to assess their possible impact on pupil progress. These analyses are among the largest and most sophisticated of their kind so far carried out for the primary phase.

Baseline assessments of over 6,000 4-year-olds made in 1994 were matched with their Key Stage 1(KS1) assessments and schools supplied data on fourteen factors about individual pupils including gender, free school meal entitlement, special educational needs, nursery education, terms in school, and for many, parents' socio-economic status. Among the eighteen school and class level factors were class organisation, class size, school type and percentage of pupils receiving free school meals. Hampshire also matched 1992 KS1 assessments for over 4,000 pupils to Key Stage 2 (KS2) assessments and collected similar background data. In future pupils will be matched from baseline to KS2 and the exercise is being repeated each

year. The relationships between prior achievement (baseline and KS1 in the two sets of analyses) and outcome (KS1 and KS2) were analysed for Maths Science and English.

Among the preliminary results for KS1 are:

- Baseline assessment in English Mathematics and Science are strongly correlated with corresponding KS1 scores (with correlation coefficients ranging from 0.56 to 0.82).
- After accounting for initial achievement at baseline, factors associated negatively with progress from baseline to KS1 are *special needs*, *free school meal entitlement*, *school absence*, and *number of primary schools attended*. Among the positive factors are *total terms in current school* and *age at time of completing KS1 assessment*. Gender is also associated with progress; girls tend to do better on reading and writing but worse in Maths and Science.
- Among school level factors, the percentage of children entitled to free school meals has a (relatively small when other factors are included) negative effect on progress. The number of adults in the class and, perhaps surprisingly, vertical grouping, have small positive effects in some subjects.
- Socio-economic group has a strong effect on progress but its effect can be accounted for by a combination of other associated variables, notably free school meal entitlement, SEN stage and level of absence.

Most significantly, schools are differentially effective in the value they add. The value added scores differ by curriculum subject and by actual baseline achievement. Some schools have high value added scores in mathematics, but not for English. Some have high value added scores in a particular subject for children with low baseline achievement but are less successful with average children. For the analyses of KS1 to KS2 scores rather similar results were obtained.

Working closely with school heads and teachers, ways have been found to convey the complex pictures revealed by these analyses to each school so that they can place the information within their local context and can discuss any problems revealed with their LEA advisers, and governing bodies. The caveats surrounding the results are also presented, notably the statistical uncertainties attached to the results so that a balanced picture can be obtained. In effect the results are treated as 'screening devices' indicating potential problems and possible remedies. The 'private' nature of the data; the fact that they are not associated with public rankings, has encouraged enthusiasm among schools (virtually all primary schools have now volunteered) and led to more openness and honesty about failures as well as successes.

Naturally, schemes such as these are capable of improvement. There is room for debate about the most appropriate analysis to carry out and important issues such as how to interpret year-to-year variation. These additional uncertainties, however, strengthen the need to regard these systems as screening devices, which are inherently imperfect, yet do provide information if used with care and understanding. In contrast to the national picture where schools are publicly ranked in misleading ways, leading often to quite inappropriate behaviour, schemes such as that in Hampshire are providing constructive and helpful feedback in a responsible manner. If government really is intent on raising 'standards' then this is the kind of scheme that offers real advantages.

References

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.

SCAA (1997). *The value added National Report*. London, Schools Curriculum and Assessment Authority.