

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

Assessing the performance of schools: limits and league tables

by

Harvey Goldstein

Institute of Education

London, WC1H 0AL

email: h.goldstein@ioe.ac.uk

Judging schools by the performances of their students.

Examination results from GCSE and A levels, National Curriculum test scores, and findings from international studies are all used to make judgements about schools and teachers. Annual 'league tables' of examination results for all schools and colleges in England and Wales are published and are intended to be used, by parents and others, as criteria for choosing among schools. In this article I shall argue that such uses have little justification. I shall argue that league tables may have a role as screening devices: that they can help to identify the relatively few institutions which (by definition) appear to have highly untypical patterns of performance and I shall discuss how this, limited, role is best realised in practice.

When discussing the use of test or examination results we should, of course, remember that educational institutions have a responsibility for encouraging children's learning and development across a much wider range of areas than reasonably can be tested

School 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. At GCSE the most prominent feature is the presentation, for each school, of the percentage who achieve 5 or more passes at grades A-C: at A-level an average grade point score is produced for each school. The national and local press are encouraged to present school and college results ranked in terms of these percentages and averages, and the Parents' Charter encourages people to use these tables in choosing schools and colleges. While the Tory government has the responsibility for introducing this system, the Labour party has provided scant criticism of it and has given little indication that it would introduce substantial amendments.

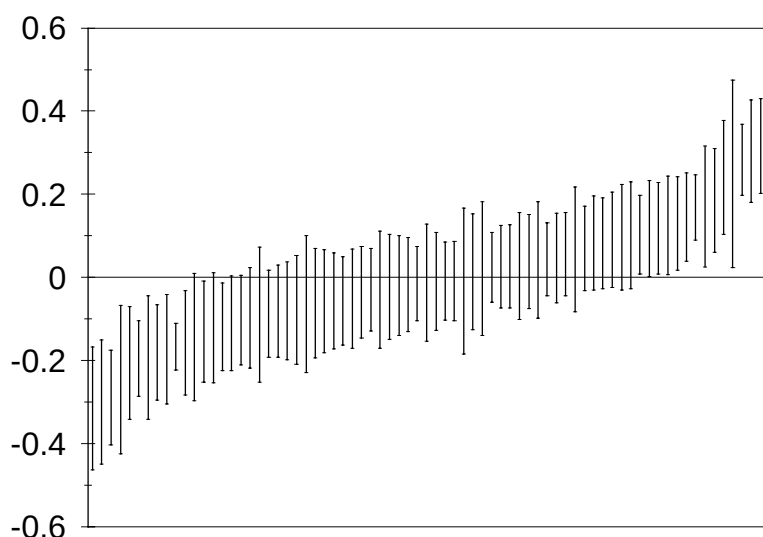
The principal argument against examination league tables is that the performance of a school is determined largely by the pre-existing achievements of the students when they enter it. Since schools differ markedly in this respect, for example some schools are highly selective, it is impossible to judge the quality of the education *within a school* solely in terms of such outputs. More recently the Government has accepted the inadequacy of using such crude rankings and has accepted the argument that what is required are so called 'value added' tables in which there is a proper allowance for pre-existing achievements (DFE, 1995): inconsistently, it

continues to promote the use of the existing unadjusted tables. There are also, however problems which apply to 'value added' tables, and I will show how initial expectations that these could provide a more sensitive indicator of school performance have failed to materialise.

The first problem with reporting only a single figure such as the overall percentage of high GCSE grades is that schools may be 'differentially effective'. Thus, for example, two schools may perform equally well on average but one may have poor performance in mathematics and good performance in English and vice versa for the other. Likewise, where value added tables are concerned, some schools may exhibit relatively good performance for initially (on intake) poorly achieving students and produce relatively weak performance for initially highly achieving students and vice versa for another school (Goldstein et al. 1993). A second problem, with both 'raw' and value added tables, is that the percentages or scores produced for each school typically have a large margin of error or 'uncertainty' associated with them. This problem is even more acute when individual subjects or departments within schools are the focus of interest, since the sometimes small numbers of students involved means that very little can be said about any individual department's performance with reasonable accuracy. In the extreme case, for some A level subjects there may be only two or three students involved and any generalisation, even over a number of years, from such small numbers is extremely hazardous.

The following figure illustrates this general problem. It is taken from a survey of some 400 schools and colleges with A level results where value added scores are calculated by adjusting for the GCSE performance of the candidates (Goldstein and Thomas 1996). The lines represent ranges of statistical uncertainty such that it is possible to judge two schools or colleges as truly having different value added scores only when the lines do not overlap. In this figure, for some three quarters of all possible comparisons of pairs of institutions, it is not possible to make such a separation. In other words, finely graded value added comparisons are of limited value since in most cases we will find no difference.

A level scores: pairwise (95%) uncertainty intervals for a random sample of schools and colleges for students in the middle (50%) GCSE score band.



Another problem with all of these tables is that inevitably they refer to a cohort of students who began their education at those institutions many years earlier. Thus, the GCSE results published in November 1996 refer to a cohort starting at their secondary schools some five years previously: given that schools can change markedly over time there will be additional uncertainty over the use of those results to predict the performance of future cohorts .

A further problem arises from recent research (Goldstein and Sammons, 1996) which shows that the primary school attended by a child exerts an important influence on GCSE performance and that this should therefore be taken into account when producing value added tables. Also, there are other factors, such as sex, ethnic origin and social class background, all of which are known to be associated with performance and progress throughout secondary schooling and which therefore will affect the interpretation of any rankings. Finally, there are several practical problems associated with producing performance tables, perhaps the most important being that during the course of a period of schooling, say from 11 to 16 years, many students will change schools. To ignore such students is likely to induce considerable biases into any comparisons, yet to include them properly would require enormous efforts at tracing them and recording their examination and test results.

Taking all these caveats together shows that attempts to rank educational institutions are fraught with difficulty. Even with extensive and good quality information, there are some inherent limitations which preclude the use of rankings other than as initial *screening instruments* to isolate possibly high or low achieving institutions or departments which can then be further investigated; bearing in mind that the information is historical. These caveats refer not only to the public presentation of comparative tables but also to the use of such information for internal purposes by individual schools as is currently being proposed by the Schools Curriculum and Assessment Authority (SCAA) for Key Stage test results where the problem of student mobility threatens to undermine the enterprise. At the very least, if comparisons among schools are to be attempted, it is very important to provide users with careful descriptions of all the limitations. If this were done it may well be that many users would find little use for league tables.

Finally, the existence of league tables within a competitive 'marketplace' has invested them with an extra importance. To have a high rank or to be labelled as improving is seen to be a competitive advantage and there will be pressure for schools and colleges to modify their behaviour to secure such an advantage. Thus, for example, a key statistic in reporting GCSE results is the percentage of subjects passed with grades A-C. By concentrating efforts on those students predicted to obtain GCSE subject grades around the C/D borderline a school may hope to increase the proportion of its grade A-C passes, but only to the detriment of relative neglect of the very low achieving or the very high achieving students. Whether intended or not, such distortions of education are hardly welcome, yet they are an inevitable consequence of such a 'high stakes' accountability system. Likewise, OFSTED inspections are required to take account of test scores and examination results, generally without proper value added information: it is perhaps not surprising that the majority of 'failures' are schools representing educationally disadvantaged populations. Even in OFSTED's own research studies, there is a

poorly understood need to take account both of intake and of the uncertainty surrounding any inferences based upon test scores and examination results (Mortimore and Goldstein, 1996).

As *one* source of information about school performance, league tables can have value, assuming that they are properly contextualised, at the very least by adjusting for intake achievement. They may indicate to LEAs for example where there are potential problems or examples of highly successful schools or departments which could usefully be followed up. They may be able to indicate, over time, where improvements or deteriorations are taking place and they can form a part of continuing research activities studying factors associated with performance. It would be unfortunate if such positive uses were to become obscured by the public promotion of league tables, value added or otherwise, as valid tools *per se* for judging schools and colleges.

Finally, despite the many abuses associated with league tables there is much to be learnt from research into institutional and system differences. We need to know more about the factors associated with success and failure of both students and institutions, but this is a painstaking, long term and complex process. Unfortunately, we appear to be passing through a phase of our culture where those in authority, or who wish to be in authority, have little taste for confronting the complexities of the real world in favour of oversimple interpretations. If such interpretations are not challenged, they may distort and degrade the systems they are supposed to support and describe.

Acknowledgements

I am most grateful to Barbara Goldstein and Kate Myers for comments on an early draft.

References

DFE (1995). GCSE to GCE A/AS value added: briefing for schools and colleges. London, Department For Education.

Goldstein, H. and S. Thomas (1996). Using examination results as indicators of school and college performance. *Journal of the Royal Statistical Society, A*, 159: 149-63.

Goldstein, H. and Sammons, P. (1996). The influence of secondary and junior schools on sixteen year examination performance. *School Effectiveness and School Improvement*, (to appear).

Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., Pan, H., Nuttall, D. and Thomas, S. (1993). A multilevel analysis of school examination results. *Oxford Review of Education*, **19**: 425-433.

Mortimore, P. and Goldstein, H. (1996). *The teaching of reading in 45 Inner London primary schools: a critical examination of OFSTED research*. London, Institute of Education. (<http://www.ioe.ac.uk/publications/ofs-crit.html>)