

PERFORMANCE INDICATORS AND LEAGUE TABLES: THEIR APPLICATION TO SCHOOLS

MEETING OF THE PARLIAMENTARY AND SCIENTIFIC COMMITTEE ON MONDAY,
29TH JANUARY 2001

Numbers can be used to inform judgments about the performance of schools, hospitals, university research departments, Government Departments, etc, but we need to use them wisely and with caution.

To what extent can league tables indicate a school's performance? Do they measure output rather than added value? What problems are associated with measuring added value rather than outputs? What confidence limits can be put on the measures? What performance indicators are there for benchmarking schools with similar problems? There is a distinction between the performance of a school and the attainment of its pupils.

League Tables and Schooling

Professor Harvey Goldstein
Institute of Education, University of London

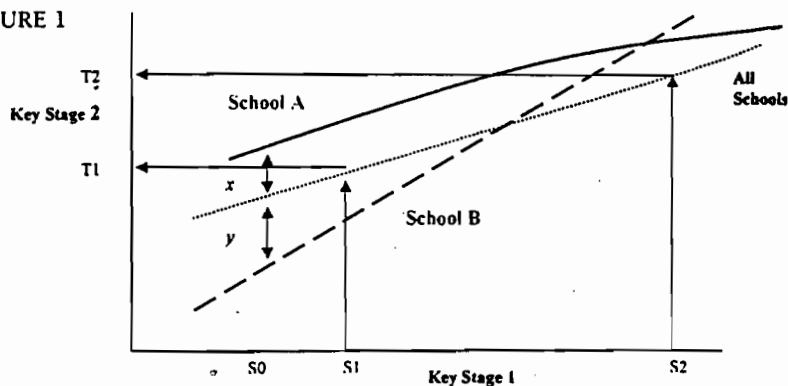


Introduction

The systematic publication of "performance tables" for key stages and exam results is now an established feature of our educational system. Here I will argue the following:

1. That the apparent simplicity of rankings of average student test and exam results is deceptive: they largely reflect "intake" achievements and, at the very least, we should adjust for intake differences – a value added approach.
2. The underlying complexity of schooling is important and requires a great deal of sophistication to understand it. Nevertheless, despite the use of complex statistical models the results of such investigations can be presented clearly and understandably. I shall refer to one LEA scheme where this happens.
3. Even with proper adjustments there are inherent limitations associated with rankings, whether in education, health or elsewhere, that should make us cautious and seek to emphasise the caveats surrounding these.
4. League tables are a poor method to ensure accountability, can distort teaching and are a poor way to measure "standards".

FIGURE 1



Value added

In figure 1 the dotted line represents the whole-country average relationship between the test scores obtained by children at key stages 1 and 2. For low scoring children at key stage 1 (score S1) the predicted key stage 2 is T1. Similarly, higher scoring children at key stage 1 (S2) have a predicted key stage 2 score of T2.

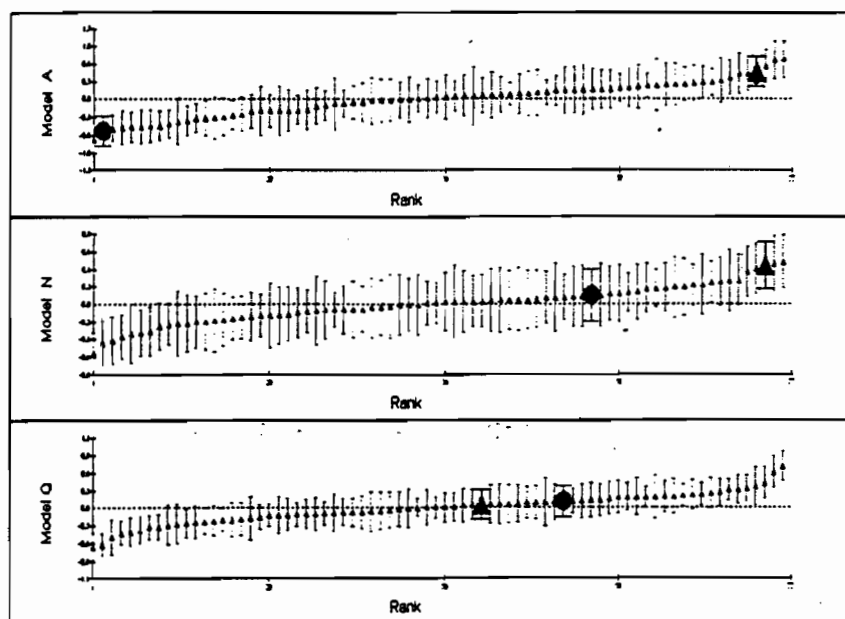
Two schools, A and B, have their own lines. At key stage 1 score S0 the predicted key stage 2 score is higher for children attending school A than for those at school B. For school A the predicted value is x above the average for the whole country and for school B y below it. x and y are the value added scores for each school at key stage 1 score S0, ie the difference between the

value predicted for a school and that predicted for all schools, at a given value of key stage 1 score. This enables us to compare "like with like". The value added will generally change with the key stage 1 score, and the situation where a school line is not parallel to the overall (dotted) line is often known as the "differential effectiveness" case. Often, the value added scores for a school are averaged over all key stage 1 scores so that a single value added score is reported. If the value added scores vary greatly across key stage 1 scores this will mask important information and is not to be recommended.

A Value added Example

Figure 2 shows how a value added approach can change the comparisons

FIGURE 2



made between schools.

Concentrating first on the top diagram of figure 2, this represents 76 schools for each of which the average score for mathematics at Key stage 2 has been calculated using data from a study in Hampshire LEA. The schools are ranked by their value added scores and for each school a (90%) uncertainty or confidence interval has been drawn: this is essentially a measure of the lack of precision or 'sampling error' attached to the average. In essence, where this interval includes the overall mean (the dotted line) there is no statistical evidence that the school average is in any way different from the overall average. Here, only about a quarter of all schools can be separated from the overall average. Thus, to quote a ranking without confidence intervals is misleading in that it implies that statements about differences can be made for all schools, whereas this is only true for a minority. Currently, none of the published educational league tables in England recognise this and that is a serious flaw.

In the second diagram the school values have now been adjusted for gender differences and whether or not the students are eligible for free school meals. This attempts to take into account social factors that may differentiate school intakes and be partly responsible for differences. We still observe wide confidence intervals. The third diagram additionally adjusts for key stage 1 scores in a full value

added analysis. Two schools have been singled out, denoted by a circle and a triangle. From both being clearly distinguishable from the overall average using just the key stage 2 scores, neither can be distinguished on a value added basis after a full value added adjustment. Note also that simply adjusting for free school meals does not give the same result as the full value added adjustment.

This hypothetical example illustrates how misleading inferences for schools can be made if either confidence intervals are omitted or value added scores are not used. Note also, that even when we use value added scores, we still have wide confidence intervals, and these represent an inherent limitation to any kind of rankings.

Constructive Uses of Performance Indicators

I have argued that there are serious limitations associated with the use of rankings of schools. Many of these arise because the "official" publication of these encourages people to take them too literally. At the very least this suggests that a task of government is to ensure that the necessary caveats about interpretation are prominent. While it has been stated government policy for some years that value added rankings are desirable, there has been little attempt to emphasise the limitations of rankings. There seems to have been little official concern with undesirable "side effects" of league tables, such as teaching to the test and curriculum

distortion. New evidence is now available from Texas about such harmful side effects and this has considerable relevance to England and Wales.

Several value added schemes have been in operation for some time. The first, pioneering, one known as ALIS uses A level data adjusted for GCSE results (and other factors) and the same team at Durham has one for primary schools. In Hampshire LEA a primary school value added project has been in existence for some years. This carries out value added analyses for baseline reception to key stage 1 and from key stage 1 to key stage 2. The results, after adjusting for a number of factors and allowing for "differential effectiveness" is fed back to schools so that each may make use of it for their own "school improvement" purposes. The results are not published in league tables and hence do not suffer from the same disadvantages as the national tables. The schools appear to be enthusiastic about them (the scheme is voluntary and almost all schools take part) and they provide information alongside other information available that may be of constructive use. The schools see the need not to overinterpret them but do find them useful.

In conclusion, I believe that enough is now understood about the construction and effects of league tables for a substantial modification in their use and presentation. Some educational systems have decided to abolish them completely (New South Wales and the Republic of Ireland for example), but at the very least their limitations need to be set out carefully and honestly so that some of the more deleterious effects can be avoided. At the same time research needs to continue into ways of obtaining better data about school performance and informative ways of presenting the results of analyses. Above all, we all need to recognise that educational reality is complex and that simple minded comparisons fail to do justice to that reality.

Note: for a fuller explanation see "The use of value added information in judging school performance" available at www.ioe.ac.uk/hg/personal/value-added-school-performance.html