

# The difficulty of ranking schools

## The limits to 'value-added'

New  
Economy

HARVEY GOLDSTEIN

*Institute of Education,  
University of London*

*A version of this article  
was previously published  
in Science in Parliament  
Whit 2001.*

The systematic publication of performance tables for the various key stages in primary and secondary schools in England, and for final exam results, is now an established feature of the educational system. However, there are a number of key issues that need to be addressed with respect to these tables:

- 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 differences in intake – a value-added approach
- the underlying complexity of schooling is important and understanding it requires a great deal of sophistication. Nevertheless, despite the use of complex statistical models, the results of such investigations can be presented clearly and comprehensibly
- 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 them
- league tables are a poor method of ensuring

accountability, can distort teaching and are a poor way of measuring standards.

### The value-added approach

The value-added approach assesses the relationship between the test scores or exam grades of a large sample of individuals at one

stage in their education with their scores at a previous stage. For example, A level results could be compared with GCSE results, or results at the end of Key Stage 2 at aged 11 could be compared with results from the end of Key Stage 1 at age 7.

A value-added score tests the difference between the value predicted for a school in terms of its test scores or exam results and that predicted for all schools, enabling comparison of like with like. Schools

can be ranked by their value-added scores and for each school a (90 per cent) uncertainty or confidence interval can be drawn: this is essentially a measure of the lack of precision, or sampling error, attached to the average. Where this interval includes the overall mean, there is no statistical evidence that the school average is

*“Misleading inferences for schools can be made either if confidence intervals are omitted or value-added scores are not used. Even when value-added scores are used, wide confidence intervals still represent an inherent limitation to any kind of ranking”*

in any way different from the overall average.

Research using data from 76 Hampshire primary schools, used to calculate the average score for mathematics at Key Stage 2, showed that only about a quarter of all schools could be separated from the overall average. Thus, quoting a ranking without the confidence intervals can be misleading in the sense that it implies that statements about differences can be made for all schools, whereas this is only true for a minority. It is a serious flaw that none of the published educational league tables in England currently recognise this.

The school values can also be adjusted for gender differences and whether or not the students are eligible for free school meals. This therefore is an additional context attempting to take into account social factors that may differentiate school intakes and be partly responsible for attainment variations. Simply adjusting for free school meals does not, however, give the same result as a full value-added adjustment.

Misleading inferences for schools can be made either if confidence intervals are omitted or value-added scores are not used. Even when value-added scores are used, there are still wide confidence intervals, and these represent an inherent limitation to any kind of ranking.

### **Constructive uses of indicators**

Many of the serious limitations associated with the use of school rankings arise because their 'official' publication encourages people to take them too literally. At the very least this suggests that a task of government is to ensure the prominence of necessary caveats about interpretation.

While it has been stated government policy for some years that value-added rankings are desirable, there has been little attempt to emphasise their limitations. There does not seem to have been much official concern with the undesirable side effects of league tables, such as teaching to the test and curriculum distortion. New evidence about such harmful side effects is now available from research

in Texas and this has considerable relevance to England and Wales (Klein *et al*, 2000).

Several value-added schemes have been in operation for some time. The first pioneering scheme, known as ALIS, uses A level data adjusted for GCSE results (and other factors) and the same team at Durham has a scheme for primary schools (Tymms *et al*, 1997).

In Hampshire 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 and they may make use of it for their own 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 are regarded as screening devices that provide information, alongside the other available information, that may be of constructive use. The schools recognise the need not to over-interpret them but do find them useful (see Goldstein *et al*, 2000).

### **In summary**

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, including New South Wales and the Republic of Ireland as well as Wales and Northern Ireland. 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 need to recognise that educational reality is complex and that simple-minded comparisons fail to do justice to that reality ●