There are serious difficulties in the path of providing reliable rankings or league tables of schools.

Assessment And Accountability

The assessment of students is now a major educational issue, whether this is done by teachers, end-of-course examinations or standardised tests. I shall describe how assessment has come to be used for accountability purposes, setting the scene with a summary of the four major roles of assessment.

1. Diagnostic or 'formative' assessment is the process whereby the knowledge and understanding of a student is assessed in order to evaluate his or her strengths and weaknesses so that both the student and the teacher can organise learning efficiently.

2. Assessment for selection purposes is the use of a test, teacher evaluation or public examination to grade students. The result of the grading is available for others to use in selection, for example for further stages of education or for employment.

It differs in function from diagnostic assessment in that the aim of the students, at the time they are assessed, is to maximise their scores or grades, rather than expose their weaknesses for subsequent correction.

This means that the format of the assessment, the attitude of the student towards it and the way it is administered will all differ in important respects from diagnostic assessment.

Both of these types of assessment are concerned with what happens to the individual student. The last two types are concerned with making statements about groups of students.

3. Assessment for institutional accountability is the use of assessment results, from the students in a school (or other institution), in order to make statements about the performance of the school. We may use standardised tests, examinations or teacher assessments to attempt to do this. The results are 'public' in the sense that institutions may be held publicly accountable for them.

4. The fourth purpose of assessment is to make comparisons at a system level — for example, between different kinds of schools or between, say, different methods for teacher reading. This purpose may be characterised as a research function, since it is trying to draw general lessons without naming individual pupils, institutions or authorities.

In the UK its most important recent manifestation was in the government-sponsored but now defunct 'Assessment of Performance Unit'.

Harvey Goldstein  
Institute of Education  
University of London

If an assessment system is to prosper and if it is to retain intellectual integrity, it must avoid claiming that it can serve conflicting aims simultaneously in terms of curriculum content, form and organisation.

Finally, it is difficult to see how the various purposes can be accommodated within a single assessment. If an assessment is to be formative, students must be encouraged to expose their weaknesses, and this is not encouraged by a selection examination; nor is it encouraged by an assessment which is designed to hold schools accountable for the performance of their students.

If an assessment system is to prosper and if it is to retain intellectual integrity, it must avoid claiming that it can serve conflicting aims simultaneously.

An important reason for the failure of National Curriculum Assessment in the early 1990s was just this advocacy of assessments serving several conflicting aims. This responsibility for this can be traced back to the Task Group on Assessment and Testing (TGAT) which reported in 1988.

The use of exam or test scores to compare schools or education authorities is practised not only in the UK but in many other educational systems. At first glance, this use of assessment for public accountability has some validity. After all, a major purpose of schools is to enhance the understanding and knowledge of their students, so that it seems quite reasonable to judge schools on this basis.

Of course, we can debate the appropriateness of different kinds of assessment, whether carried out by teachers or, for example, by rigorously applied standardised tests. We can also have a useful debate about exactly what is to be assessed for this purpose. In what follows I shall concentrate on the general validity of any assessment designed for accountability purposes.

The first, perhaps obvious point to be made is that the knowledge and understanding that students attain is determined not only by what they experience at school, but also by their experiences before they entered any particular school, in their homes, in their communities, and also by a result of their biological or genetic make-up.

Thus, if we compare two secondary schools whose intakes differ considerably in terms of the pre-existing performance of their students, we would expect the students in the school with the higher-performing intake to perform better on any subsequent assessment. We expect this because we realise that the other school would have to work harder to achieve the same result.

On the other hand, if the students at each school were very similar in
It is important to avoid the trap of supposing that the provision of some information about schools, even just unadjusted raw league tables, is better than no information at all.

The possibility of differential value-added performance. Whether we have carried out a value-added analysis or even a simple unadjusted analysis, all comparisons between schools will be subject to a margin of statistical 'error' or 'uncertainty'. This will be relatively large when the number of students in a school is small.

For example, the Newcastle ALIS scheme that carries out value-added analyses of A-level results reports comparisons on a subject-department basis where numbers may be very small indeed.

Work at the London Institute of Education has shown that for the GCSE examination, value-added analyses yield 'uncertainty intervals' which are so large that only schools with extremely small or extremely large value-added scores can be separated from the remainder. In other words, it is not technically possible with any reasonable level of certainty to give an unequivocal ranking of schools in a league table, whether this is an unadjusted 'raw' table or an adjusted value-added one.

Finally, besides the need to take account of intake assessment, it may also be important to take account of other factors, outside the general records for individual students to be linked together over their school careers — an expensive and time-consuming activity.

There is one further, fundamental problem with using assessments to make judgements about schools: namely, that such judgements can be seriously out of date. For example, if we were to use GCSE results for the 1994 cohort of students (whether value-added or not) to judge a school's performance, this would refer to a group of students who had entered their schools five years previously in 1989.

Because schools can change markedly over such a period, the use of 1994 results to predict the performance of a cohort who entered school in 1994 and whose exams were in 1999, is likely to be a perilous business.

I am not against value-added analysis, I think it is important to carry out such analyses for research purposes in order to study the factors that are responsible for some schools appearing more effective than others for particular groups of students. My doubts are whether the results of such analyses can be used to hold schools accountable.

If there is a political wish to find ways of comparing how schools set about their tasks, then there are other perfectly proper and direct methods for so doing.

control of the school, which may be responsible for influencing performance. Home environment is clearly one of these, but so are other factors such as resources available for teaching etc.

I hope that this brief outline of some key issues has made it clear that there are serious difficulties in the path of providing reliable rankings or league tables of schools. Not only are comparisons that are based upon raw results misleading and potentially unfair, we should not expect even value-added analyses to provide definitive comparisons. Furthermore, value-added analyses, although they are certainly an improvement, are necessarily more complicated to carry out. They require intake assessments, and

If there is a political wish to find ways of comparing how schools set about their tasks, then there are other perfectly proper and direct methods for so doing.