J. EDUCATION POLICY, 1988, VOL. 3, NO. 2, 197–202

A note on national assessment and school comparisons

Harvey Goldstein
Institute of Education, University of London
Peter Cuttance
University of Edinburgh

Introduction

The Government's Task Group on Assessment and Testing have produced a lengthy report (Black 1988) which, if implemented, will introduce major changes both in teaching and in the relationships among schools, parents and LEAs. Closely linked to the proposed National Curriculum, it makes proposals for measuring 'the delivery of the national curriculum in most schools'. It suggests an extensive system of assessment at 7, 11, 14 and 16 years of age, to consist of centralized tests and teacher judgements, with the latter brought into line with the former through a system of 'moderation'. At the latter three ages it proposes that the distribution of results for each school is published so that schools can be compared, for example in terms of averages or, say, proportions of pupils with 'low' scores.

The report covers an extremely wide range of issues and makes a large number of recommendations, many of which will no doubt be debated in the next few months. In this article we focus on the one aspect of publication of school results; for two reasons. First, the report itself spends relatively little time discussing the real problems this would raise, and second, in our view, such publication could well become the most visible and perhaps the most important aspect of the whole system, irrespective of the actual content of the assessments or indeed of the National Curriculum itself.

Reporting school results

Over the last few years a number of LEAs as well as the DES have become interested in the reporting of average examination results for schools and LEAs. Recent research has, however, shown that such *league tables* can be highly misleading (Goldstein and Woodhouse 1988). As the Task Group report itself says, 'a school's performance can only be fairly judged by taking account of many aspects of its work and of many factors outside its control that affect its work'. A vast research literature documents the fact that children from socially disadvantaged areas tend to have lower exam scores and test results than those from more socially advantaged areas (cf. Coleman *et al.* 1966, Garner 1987, Madaus *et al.* 1980, Plowden 1966, Rutter *et al.* 1979). To compare simple school averages or distributions would in part reflect such differences and obscure any real 'effects' due to the schools themselves.

Perhaps most important of all, research in this area has demonstrated that the attainment of the children when they first enter the school is the single most important determinant of subsequent achievement, and a growing literature exists both documenting this and discussing how fair and valid school comparisons can be made by taking it into account

(Cuttance 1988a, Gray *et al.* 1986, Raudenbush and Bryk 1986, Raudenbush and Willms 1988).

The Task Group report does not mention this research on *school effectiveness*; nor does it acknowledge the importance of allowing for pupil 'intake'. On the general issue of whether to adjust for external factors, such as social class, it considers that to do so 'would be liable to lead to complacency', while not to do so would lead to 'misinterpretation'. Its solution is that aggregate school results should be published along with 'a general report for the area . . . to indicate the nature of socio-economic and other influences which are known to affect schools'. In effect the report is saying that the aggregated results, which will create 'confidence in the measurement precision' are to be evaluated using a further set of measurements and descriptions of factors which might affect those results. In essence, it is not really suggesting that adjustments should be avoided; rather it is passing on the responsibility for making them to others, notably parents and LEAs. In the remainder of this article we look at this problem of school comparisons in more detail to see what may or may not be sensible.

Comparing school averages

There are major difficulties in using *average results* to compare schools, whether these be based on exam results or more sensitive teacher evaluations, and these greatly outweigh the advantages of simplicity and cost. Whether the 'raw' unadjusted results are used or the results adjusted for social background and intake, there exist extremely difficult problems of interpretation, because the estimates of school performance, particularly their relative ranking, are sensitive to the method used.

At one extreme, the more background factors which are taken into account, the more unstable and unreliable become the resulting comparisons. Recent work carried out at the London Institute of Education, for example, has shown that existing *league tables* of LEAs based on average exam results, can have their rankings changed markedly, and haphazardly, by making trivial modifications to the adjustment procedures (Goldstein and Woodhouse 1988). Furthermore, a comparison of school averages, or distributions, tells us nothing about the relative achievements of different types of pupils within the schools. Two schools may achieve the same average result by quite different means. Thus, one school may obtain good results with pupils whose intake achievements are low, but relatively poor results with pupils of high intake achievement, whereas a second school may produce homogeneous results for all pupils. Consequently schools which perform well relative to other schools for the average pupil in the population may perform less well for disadvantaged or advantaged pupils (Cuttance 1988a, Raudenbush and Bryk 1986).

Analysing pupil achievement

A more appropriate way to approach comparisons of schools is via the analysis of results for individual pupils, taking into account intake and other appropriate background factors. This January, in London, the first International Conference on Effective Schools was held, and it was clear that there is a great deal of interesting work being pursued by groups in a number of countries. The recent development of new *multilevel* data-analysis techniques, ie, techniques that recognize that the data on pupil attainment carry information on influences at several levels (pupil-classroom-school-LEA) (Aitkin and Longford 1986, Goldstein 1987, Raudenbush and Bryk 1986), has sustained much of this work, although general agreement

is yet to be reached on some issues. Outstanding issues relate to the characteristics of pupils that should be taken into account at entry to school (Cuttance 1988 b) and statistical problems in estimating the performance of schools in the absence of information on the causes of school effects on pupil attainment (Raudenbush and Bryk 1987, Raudenbush and Willms 1988).

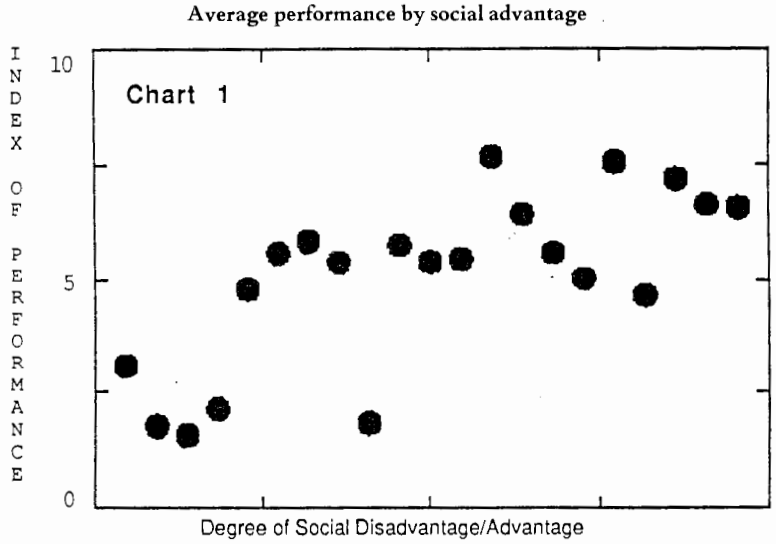
It is clear from the Task Group's report that they have not contemplated the comparison of schools by any method other than by the simple aggregation of results across all pupils in a school. In our view such a procedure would not only be a waste of time and money, but could also lead to unfair judgements that would have undesirable consequences.

An example of the problems in comparing schools

We can illustrate some of the problems of comparing schools with the method recommended in the Report by using data for secondary schools from one LEA.[1] The symbols in chart 1 show the average level of attainment of pupils in each of the 21 schools in one LEA. The method recommended in the Report is for such aggregate measures of school performance to be presented in the context of the socio-economic characteristics of the community that the school serves. Hence, we rank order the schools from left to right according to the degree of social disadvantage or advantage in their catchment area, based on Census data, with the more disadvantaged schools to the left of the chart.[2]

Two points are clear from chart 1. First, there is apparently substantial variation in the performance of schools in this LEA. But we will soon see that this is a totally misleading description of the variation in the effectiveness of these schools. The other point to note from the chart is that schools in the more disadvantaged communities appear to perform at a lower level than those in the socially more advantaged communities.

The assessment of the performance of schools as presented in this chart is broadly in line with the way that the method recommended in the Report would be implemented in many LEAs. This assessment takes account of the variation in the socio-economic characteristics of communities, but it makes no adjustment for the prior attainment of individual pupils entering schools, nor for characteristics of their social background.



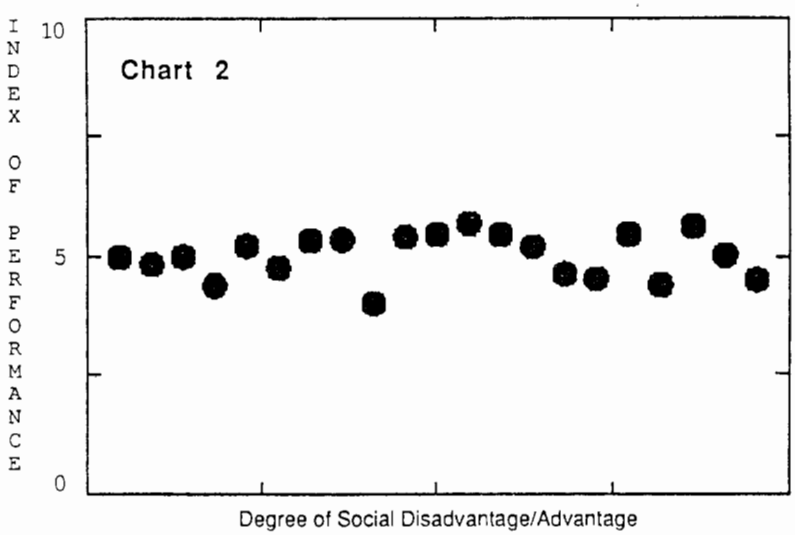
The Educational Reform Bill 1988 makes provision for greater movement among schools via its revisions to the Parent's Charter, and the relationship between the socio-economic characteristics of local communities and their schools is likely to become weaker than it is at present. This will make comparisons of schools based on aggregate catchment data more difficult in the future than it is at present, because the location of a school will not necessarily correspond to the pattern of residential distribution among it's pupils. Measures of *individual-pupil* social background explain much more of the variation in pupil attainment than *aggregate* measures of community socio-economic characteristics. But the big gains in explaining pupil attainment come from the incorporation of measures of pupil prior attainment, in addition to social background, into models. Together, these two sets of factors explain over half of the variation in pupil-level attainment.

A much more accurate assessment of the performance of schools is therefore obtained by taking account of these characteristics of pupils in school intakes. The basic task in assessing the effectiveness of schools is one of estimating the *adjusted* attainment of individual pupils after taking account of the prior attainments and social characteristics of pupils. The variation in the intakes to schools operates as a sort of handicapping system. The schools with the largest handicap are those whose pupils have the lowest levels of prior attainment and social advantage.

Chart 2 shows the estimates of school performance that are obtained when this handicapping system is taken into account by adjusting attainment at pupil-level, using the multilevel modelling technique described in the appendix. Compared to the earlier chart these results are striking for two reasons. First, they show differences between schools which are much smaller than those suggested by the aggregate measures of unadjusted pupil attainment in chart 1, although the differences are still educationally significant.[3] Second, they indicate that the schools in socially disadvantaged areas in this LEA are not performing poorly, given the prior attainments of the pupils entering them. The results indeed suggest that some of the socially more advantaged schools in the LEA are the ones that are underperforming, given their intakes.

This does not mean that the pupils in the socially disadvantaged and the socially advantaged schools will leave school with the same absolute levels of attainment. The pupils in the

Performance by social advantage after accounting for 'intake' in a multilevel analysis



less advantaged schools will leave school with a relatively lower level of qualifications than those from the socially advantaged schools, because they entered their secondary schooling with lower levels of attainment at the completion of their primary schooling.

But the appropriate measure of the performance of schools is not the absolute level of attainment of their pupils, rather it is the amount of progress that pupils make while attending schools. To fail to take full account of the pupil intakes to schools is equivalent to assessing the performance of businesses without taking account of the cost of their raw materials.

As the two charts clearly show, a failure to take account of the intakes to schools at the pupil-level, rather than informally through a report on the socio-economic characteristics of the school's community, is a fundamental flaw in the Report's proposed system of school evaluation. In any case, the right of parents to choose schools under the Parent's Charter will make any informal adjustments based on census data for the local community quite unreliable by the time the national assessment system is introduced. The information gained from such a system may have no clear relationship to the effectiveness of schools, and is likely to result in the unjustified victimization of schools in socially disadvantaged communities, while failing to locate the poorly performing schools in the socially advantaged communities. In short, the system of school evaluation proposed in the Report would be quite misleading, and school boards, parents, and teachers would be right to protest loudly at the use of such poorly constructed performance indicators.

In their future considerations we would urge the Task Group, and the Government, to give serious attention of these issues. They should envisage that without a substantial rethink, and the will to invest in the best developments of current research on the effectiveness of schooling, the best course of action might be to drop the idea of using the results of the national assessments of pupils to compare schools.

Notes

1. The work on which this article is based was funded in part by grants from the ESRC and from the Scottish Education Department to the Centre for Educational Sociology (CES) at the University of Edinburgh where a programme of research on the effectiveness of schooling has been underway since the early 1980s. The views expressed are, however, solely those of the authors. We should like to thank colleagues at the CES for their preparation of the data on which the examples are based.

2. See the technical appendix for a description of the data employed in these examples. These data are employed as a basis for showing the differences in the two models for a particular data set. The variation in the estimates for the two models will, in general, be data dependent, thus the estimates for this particular data should not be read as if they were directly generalisable to other LEAs. They serve merely to show that the models provide different impressions of the effectiveness of schools in this one LEA. Aitkin and Longford (1986) show that these two models provide different impressions of the effectiveness of schools in another LEA also.

3. The measures of social disadvantage and advantage are based upon estimates of the degree of social deprivation in the areas served by schools. Garner (1988) provides details of the construction of the index of deprivation from Census data and of its use in this context. We should like to thank Cathy Garner for providing us with the data relating to the extent of social deprivation in these school catchments.

4. The magnitude of the differences in performance among schools are discussed in detail in Cuttance (1988 a), where they are related to the magnitude of the change in general levels of performance over time.

References

AITKIN, M. and LONGFORD, N. (1986) 'Statistical modelling issues in school effectiveness studies', *Journal of the Royal Statistical Society*, Series A, 144, pp.419–461.

BLACK, P. J. (Chairman) (1988) *National Curriculum: Report of the Task Group on Assessment and Testing* (London: Department of Education and Science).

BURNHILL, P., LAMB, J., and WESTON, P. (1984) *Collaborative Research Dictionary and Questionnaires 1981* (Edinburgh: Centre for Educational Sociology, University of Edunburgh).

CUTTANCE, P. F. (1988 a) 'Modelling variation in the effectiveness of schooling', (Edinburgh: Centre for Educational Sociology, University of Edinburgh).

CUTTANCE, P. F. (1988 b) 'Assessing the effectiveness of schools', (Edinburgh: Centre for Educational Sociology, University of Edinburgh).

DES (1967) *Children and their primary schools*, The Plowden Report (London HMSO).

GARNER, C. L. (1988) 'Educational attainment in Glasgow: the role of neighbourhood deprivation', in E. Bondi and M. H. Matthews (eds.) *Education and Society* (London: Croom Helm).

GOLDSTEIN, H. (1987) *Multilevel Models in Educational and Social Research* (London: Charles Griffin & Co).

GOLDSTEIN, H. and WOODHOUSE, G. (1988) 'Educational performance indicators and LEA league tables.' (London: Institute of Education, University of London).

GRAY, J., JESSON, D. and JONES, B. (1986) 'The search for a fairer way of comparing schools' examination results', *Research Papers in Education*, 1, pp.91–119.

RAUDENBUSH, S. and BRYK, A. (1986) 'A hierarchical model for studying school effects', *Sociology of Education*, 59, pp.1–17.

RAUDENBUSH, S. and BRYK, A. (1987) 'Quantitative models for estimating teacher and school effectiveness'. Paper presented at the Invitational Conference of the Center for Student Testing, Evaluation, and Standards, Princeton, New Jersey.

RAUDENBUSH, S. and WILLMS, J. D. (1988) 'Procedures for reducing the bias in the estimation of school effects.' Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, 5–9 April.

Technical appendix

The data employed for estimating the performance of schools in charts 1 and 2 is drawn from the 1981 Scottish National School Leavers Survey (Burnhill, Lamb, and Weston 1984). The data is for the 1521 leavers from 21 schools in one local authority. The attainment measure employed is a scaled score based on the pupil's fourth year Scottish Certificate of Education Ordinary Grade examination result. Pupils who did not enter the exam were assumed to have done so because they were unlikely to pass, except in the few cases where they bypassed directly to the fifth year syllabus. The details of the scoring and other characteristics of the attainment measure are provided in Cuttance (1988 a). Further details of the data are available on request from Peter Cuttance.

A score of 5 on the index of performance in charts 1 and 2 corresponds to the national average pupil-level attainment score. A score of 2·5 represents one standard deviation below the mean in the distribution of pupil attainment, and a score of 7·5 is one standard deviation above the mean in this pupil-level distribution.

The performance index in chart 1 is based on a simple average of the pupil attainment score within each school.

Chart 2 is based on a more complex multilevel statistical model. This model controls for social class, mother's education, number of siblings in the family, gender, and pupil ability measured in primary 7. It incorporates a between-school and a within-school component of variation in attainment, to give a 2-level model (see Goldstein, 1987, chapter 2).