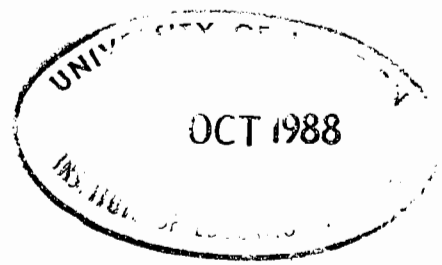
National Assessment and Testing: A Research Response

Papers presented to the
British Educational Research Association Conference on
'Benchmark Testing' held at
The London Institute of Education on 11 February 1988

Edited by **Harry Torrance**

Contributors **Paul Black, Patricia Broadfoot,
Margaret Brown, Sally Brown
Harvey Goldstein, Roger Murphy**



Published by the
British Educational Research Association
April 1988

Comparing Schools

Harvey Goldstein

Introduction

In this contribution I shall concentrate on a single, but key, aspect of the TGAT report, namely the publication of assessment results at the school level. It is clear that an 'objective' quantitative basis for ranking the achievements of schools is intended to facilitate other aspects of the education bill. Thus, for example, parents would be able to base school choice on such published results. It is important, therefore, to examine the basis for such reporting and the problems it would raise. I will be concerned with two related issues. First the validity of using aggregate level results, and secondly the problem of making due allowance for factors which might influence these results.

School Level Aggregation

The TGAT report is quite clear that at ages 11, 14, 16 and possibly at 7 also, assessment results, aggregated to school level, should be published. The report advocates the reporting of a school's distribution of results. Nevertheless, whether it is a simple mean or, for example, the proportion of pupils with high grades or rankings, the same general problems are apparent.

It is clear also from the report that a school's results effectively will consist of its results on the nationally standardised assessments. Apart from 'particular cases' the intention is that the training of teachers in assessment should have a principal, but narrow, aim of making their own assessments conform to the national test score or grade distributions (para 62). Thus, aside from the highly questionable nature of such an undertaking, the teacher assessments will be strictly superfluous when school aggregate results are computed, since these assessments already will have been made to agree with the score or grade distributions on the national assessments.

There are two serious drawbacks to the use of aggregate level results. The report itself recognises that just to publish school averages (or distributions) is misleading. It realises that these should be contextualised and notes that "a school's performance can only be fairly judged by taking account of many aspects of its work". The report's solution is that aggregated results should be published along with "a general report for the area, . . . to indicate the nature of socioeconomic and other influences which are known to affect schools".

In effect, this proposal gives responsibility for taking account of such factors to the potential user; the parent or LEA for example. This supposes, however, that judging school performance in the light of socioeconomic and other circumstances is a relatively straightforward matter. In fact this is far from true and attempts by researchers to do just this have been notable for their lack of success. In particular, Woodhouse and Goldstein (1988) show that in the case of the LEA aggregate level comparisons, making allowance for influential factors leads to inherent problems and allows no consistent interpretations. They show that minor changes to measurement scales etc., produce large changes in the resulting rankings of LEA's, and they conclude that analyses of aggregated data are inherently unreliable. It is hardly conceivable that the procedure advocated in the TGAT report will produce anything

better, based as it is upon informal, subjective, interpretations. If, despite this, the suggested procedures are implemented, it would soon become clear that a variety of interpretations was possible, with no objective resolution. In these circumstances there would be a real danger that socioeconomic and other factors simply would be ignored and the crudest kind of 'league tables' would result.

The second problem with the TGAT recommendations is that there is no serious recognition of the importance of allowing for the achievement of pupils when they enter schools. There is just a passing reference (para 138) to a secondary school "at its discretion" referring to its 11 year entry assessment results when reporting age 14 results.

It is now recognised by those working in the area of 'school effectiveness' studies (Royal Statistical Society, 1984) that 'intake' achievement is the single most important factor affecting subsequent achievement, and that the only fair way to compare schools is on the basis of how much ***progress*** pupils make during their time at school. Nowhere in the TGAT report is there a reference to this body of research knowledge.

Multilevel Analysis

Following the work of Aitkin and Longford (1986) it is now widely accepted that the only potentially valid method for comparing institutional effects is via the simultaneous modelling of data at all the relevant 'levels'. In the case of schools this means that results for individual pupils should be related to those pupils' other characteristics, including 'intake' achievement, together with relevant school characteristics such as the socioeconomic character of the catchment area. These relationships can be used to predict a school's performance and this predicted performance then compared with the actual one. The difference between these two for each school, its residual, can then be used as the basis for a measure of school effectiveness. In particular, schools can be compared in terms of the performance of particular kinds of pupils. Thus, for example, one school may produce relatively good results (in terms of progress) for pupils with low intake achievements, but not for those with high intake achievements. Another school might produce homogeneous results for all pupils. Despite such differences, the ***average*** results for the schools could be the same, in which case the interesting detailed effects would be concealed by an analysis based solely on school averages. It is just such interesting details which would seem to be of most interest to parents, LEA's etc.

It must be said that there are many problems surrounding the interpretation of such multilevel analyses (see e.g. Goldstein, 1987) and this is now an active area for research worldwide. One particularly troublesome issue concerns the scaling of assessment results. Typically, whether they are exam results or, say, profile summaries, the scales or cut-off points essentially are arbitrary. Thus, 16+ public exam results are often summarised using a rating scale first adopted by the ILEA (Aitkin and Longford, 1986) which gives a numerical score to each grade and then averages these scores. In general, a different scoring system would give different final rankings, not only of pupils, but also of schools. This will be true for all such summary scores or gradings and the indeterminancy associated with any rankings will apply whether the TGAT proposals are adopted or whether a multilevel model is used. An empirical demonstration of this is given by Goldstein (1987). Another issue is concerned with the year to year variation in performance of schools. As yet, there is little evidence about how much of the residual variation between schools can be ascribed to each school's own year to year variation. If the latter variation is relatively large then there

will be only a small correlation between school rankings from one year to the next and this would cause obvious problems for potential users.

The statistical methodology is still being developed and there is not yet any large scale implementation which would allow a thorough study of such practical and theoretical problems.

Conclusions

I have argued that the TGAT proposals for school comparisons are poorly conceived and ignore relevant research. If they are implemented in their present form, the results will be both misleading and unfair. Moreover, any system based solely upon aggregate results will suffer from the same problems. If the Government seriously wishes to provide informative school comparisons, it must, at the very least, look to the use of multilevel procedures. If it were prepared to do this, and given the novelty of such procedures, it would also have to be prepared to set up suitable feasibility studies before any widespread implementation.

In any case, whether or not the Government decides to invest in a more productive approach, it should recognise the inadequacy of the present proposals and should not proceed with their implementation.

References

Aitkin M, and Longford N. (1986). Statistical Modelling in School Effectiveness studies (with discussion). *J. Royal Statistical Society*. A, 149, 143.

Goldstein H. (1987). *Multilevel Models in Educational and Social Research*. London, Griffin; New, York, Oxford University Press.

Royal Statistical Society (1984) Assessment of Examination performance in different types of schools. (A general discussion). *J. Royal Statistical Society*, A. 147, 56981

Woodhouse, G and Goldstein H. (1988). Educational Performance Indicators and Local Authority League tables. Submitted for Publication.

Benchmark Testing: Scottish Developments

Sally Brown

The Government's decisions on assessment and testing in Scotland seem to be lagging behind those in England and Wales. We received the Consultation Paper ***Curriculum and Assessment in Scotland: A Policy for the 90s*** in November 1987 and comments on the Secretary of State's proposals are invited to arrive by 26 February 1988.

I say '***seem*** to be lagging' because final decisions on educational matters can move very quickly north of the border, and there is no indication that the Secretary of State for Scotland will follow the Rt. Hon. Kenneth Baker's lead by setting up a Task Group to advise him on assessment and testing. What he has done is to commission before the end of the consultation period two short projects (to be carried out by the Scottish Council for Research in Education and the University of Edinburgh) to prepare for the Scottish Education Department (SED) reports for ***internal*** use. No external dissemination is intended and copyright will be retained by the SED.

Since the focus of these papers is on assessment, I do not wish to spend time comparing what is proposed for the 'national curriculum' in Scotland and that for England and Wales. There are, however, a few points which should be made about the Scottish scene if the similarities and differences between the two contexts are to be appreciated.

First, the SED is not directing its attention to the curriculum and assessment of 14 to 18 year olds; it asserts that for them 'the development of Standard Grade and the Action Plan . . . will offer a substantially improved curriculum and much better assessment arrangements'. Because Scotland does not have multiple examination boards and other awarding bodies, the control of the curriculum and assessment, in the later years of secondary school and in further education, is effectively maintained through the Scottish Examination Board (SEB) and the Scottish Vocational Education Council (SCOTVEC).

Secondly, in its concentration on education from 5 to 14, the consultative paper looks for 'clearer definition than at present of the content and objectives of the curriculum'. It states that 10 to 12 year olds (Primary 6 and Primary 7) 'are insufficiently challenged . . . (and) there is need for more progression and rigour in the primary curriculum'. The importance of a curriculum which meets 'the needs of the less able so that their potential can be developed fully'is accepted, and attention is drawn to 'a serious problem of curricular discontinuity, especially in the four years between P6 and S2'.

Thirdly, the framework for the secondary school curriculum provided by the Consultative Committee on Curriculum (CCC) is endorsed. This is subject-based and ensures that secondary pupils' programmes all include language, mathematics, science, social studies, technological activities, art, music, or drama, religious and moral education and physical education. The CCC's description of the primary curriculum is also accepted i.e. language arts, mathematics, environmental studies,