

Standards of Research

Harvey Goldstein

Professor of statistics at the Institute of Education, University of London, Harvey Goldstein was a member of the DES advisory group overseeing the National Children's Bureau report comparing children's progress in selective and non-selective schools. Here he looks at another study comparing such schools, **Standards in English Schools** by Marks, Cox and Pomian-Srzednicki, published by the National Council for Educational Standards, and finds it seriously flawed.

Ever since the early Black Paper days of the 1960s, opponents of comprehensive school reorganisation have been assembling statistics to show that children in grammar schools perform better at examinations than those in comprehensives.

Their early attempts were strongly criticised for ignoring the fact that the selective intake of the grammar schools virtually guaranteed better exam results, and that a fair comparison could be carried out only if proper allowance was made for school 'intake' characteristics. These criticisms seem to have been accepted and by the 1977 Black Paper, Baldwin was attempting to incorporate such allowances. Like others, however, his analysis was flawed from a lack of appropriate 'intake' measurements on the individual children in the study.

In 1980, the National Children's Bureau (NCB) entered the debate with its own study (**Progress in Secondary Schools**, by Jane Steedman), which for the first time was able to make appropriate allowances for intake differences using longitudinal information from a large cohort study, and showed few important differences between selective and non-selective schools.

Following this, it was the turn of Caroline Cox and John Marks, in the Black Paper tradition but writing from the Centre for Policy Studies, to attack the NCB study (see **Forum** Vol.23, No.3, 1981, for an account). This attack, interestingly enough, was not aimed at the innovatory approach of using intake score adjustments, but rather at the researchers themselves, who were accused of political 'bias' and 'secretiveness'. In addition Cox and Marks attacked the use of test scores (rather than exam results) claiming that these were inadequate in several ways. The strident tone and intellectual weakness of these criticisms was such as to induce the British Educational Research Association to issue a public rebuke to these critics (**BERA Newsletter**, April 1981). More recently, the NCB has supplemented its original findings by substituting exam results for test scores (**Exam Results in Selective and Non-selective Schools**, by Jane Steedman), and concluded once more that there were few important differences between selective and non-selective schools.

Hard on the heels of this NCB report now arrives the report by Marks, Cox and Pomian-Srzednicki (MCP), this time under the imprint of the National Council for Educational Standards (NCES). While both studies are concerned with exam results, the NCB study was able to make allowances for the intake characteristics of the children in the study, because it was longitudinal. The

NCES study on the other hand had to content itself with allowances at LEA level only, ignoring any variation within LEAs. Potentially, therefore the NCB study was able to carry out much more satisfactory analyses than the NCES one. Both of the NCB studies provide far and away the best available evidence for comparing selective and non-selective schools. Despite the imperfections of these two studies, which are discussed in the NCB report, it is extremely unlikely that anyone could obtain data approaching their quality. This was the view held by the advisory committee set up by the Secretary of State when it reported to him on the first NCB study and it was also their hope that the existence of the NCB study would discourage, inevitably inferior, collection and analysis of other data. From such a perspective, therefore, the NCES study hardly required a detailed critique since its basic design was bound to yield inferior conclusions to those of the NCB. Indeed, it is somewhat curious, after the fuss made by Cox and Marks over the availability of data from the first NCB study, that they did not spend their energies re-analysing the NCB data. Nevertheless, since the NCES study does exist and since an analysis of their methods is instructive, I shall discuss their methods and conclusions.

The NCES Report

The NCES study took advantage of the 1980 Education Act which requires schools to make their public examination results available. MCP wrote to LEAs and individual schools requesting 1981 exam result details for all fifth formers. They had responses from 55 per cent of LEAs and 49 per cent of secondary schools and also used published data about LEA social class composition, ethnic mix, expenditure per pupil and pupil/teacher ratio. In addition they classified each school as comprehensive, grammar or secondary modern.

The principal analyses compared types of school, allowing for social class differences, and studies variation between local authorities with particular reference to the relationship between expenditure patterns and exam results.

The reason for allowing, or adjusting, for social class, is that it is associated with educational attainments and separately, as MCP themselves find, with the degree of comprehensiveness of an LEA, the more working class LEAs tending to have greater proportions of comprehensive schools. Thus, in order to attempt any kind of cause-effect inference, at the very least a

researcher should take this into account. The NCES study classified LEAs into three groups according to their social class composition.

Two difficulties arise immediately. First, the classification is very coarse, being only the proportion of children in social classes four and five, and almost certainly does not account for the full social class effect (see later). Secondly, social class is measured at LEA level, whereas there is known to be considerable variation within any LEA in terms of school catchment area social class. Thus, even if the report were to take this analysis as far as is theoretically possible, by comparing school types within each LEA, it is obvious that any remaining association between school type and exam results could well be due to social class variations within LEAs — in addition of course to variations in intake characteristics about which the NCES study has no information whatsoever.

Nor is this mere speculation. One authority, the ILEA, for several years now has carried out analyses in which it compares school exam results after adjusting for average school verbal reasoning score at intake and a number of social factors, (**School Examination results in the ILEA, 1976-1982**, research reports, ILEA Research and Statistics Group). In the 1978 report it was shown that a correlation of 0.98 could be achieved between a suitable combination of intake and social factors, and examination results. Gray and Jones (**TES**, 15 July 1983) state that in their own study correlations of 0.8 are readily obtainable. In other words it clearly is possible to characterise schools so that differences in average exam performance become highly predictable, leaving the possibility for only a negligible remaining differentiation in terms of school type.

Hence, comparisons of school types at the level of the school, if carried out as accurately as possible, seem destined to be extremely uninformative about school type differences. This implies that it is only analyses at the pupil level which are able to provide useful comparisons. Indeed, it hardly needs saying that the effects on different types of pupils are of more interest than the average effect for a school and of course it is precisely in this respect that the NCB study is strong. This is not to deny that school level studies have other merits. For example, it would be useful to replicate the ILEA analyses, in other LEAs, and the publication of exam results by schools requires research to explore ways in which knowledge of the results might be useful to parents and others.

As far as the between-LEA comparisons are concerned, the NCES report claims that the average LEA exam results show wide variation even after adjusting for social class, and that increased spending per pupil does not necessarily lead to better exam results.

In the first case, as Gray and Jones pointed out, the coarseness of the social class grouping is a problem and when a finer classification is used by them the variability between LEAs accounted for by social class goes up from the 30 per cent of the NCES to 70 per cent. Just as with schools there should be little difficulty in predicting LEA average exam results with great accuracy given suitable measurements.

On the question of spending, the NCES report finds that the higher the expenditure per pupil the poorer the average LEA exam result. One obvious explanation for

this would be that the LEAs who spend more tend to be those with a greater proportion of educational problems such as social disadvantage which are linked to poorer exam results and also which require higher spending, a point long recognised by the DES in its educational spending allocations. The NCES report finds that this negative relationship exists after making allowance for social class and other factors, but the same problem about coarseness of measurements applies. In fact, the report itself urges caution, saying that these results are 'not as clear cut and consistent' and that 'it may be that unambiguous answers to the intriguing questions raised by the preliminary analyses cannot be given unless more data are collected'. Unfortunately, by the time the reader reaches the summary at the back of the report this has been transmuted into 'some of our analyses suggest that providing more teachers and spending more money per pupil does not necessarily lead to better exam results'!

The NCES study has had considerable publicity; more so than the technically superior NCB study. In part this may be due to the NCB study appearing in the middle of the 1983 general election campaign and largely being over-looked. In part it may be due to more efficient publicity by the NCES and in part it may be that the NCES conclusions are more acceptable to the majority of the media than those of the NCB. Whatever the reason, the NCES intend to produce further reports, including it seems, studies of individual LEAs. In the present climate, unhappily, their statements are likely to be given favourable publicity, whatever their intellectual content. However, the recent disclosure that the DES statisticians, in a report to the Secretary of State, have been severely critical of the NCES study, may reduce credibility of the NCES (**Guardian**, 1 October 1983).

Given their existing record it would be safe to maintain a scepticism about any future reports from the NCES, although one must always allow that the quality may improve. This might occur if the NCES were actively to encourage other researchers to analyse their data and, more importantly, if they were to invite active workers in this area to comment on their analyses before final publication. If the NCES wishes to be regarded as a competent research body, then it should submit its research to the same discipline as do other researchers, with recognised standards of analysis and presentation and well tried methods of debate.

Public arguments over controversial research findings, I suspect, leave most lay persons confused and inclined to declare a curse on both houses. Consequently, standards of education debate suffer and both educational research and education itself become the poorer. Needless to say, detailed scrutiny by other researchers is no guarantee of quality, but it should help to produce more balanced and less error-prone results, especially if the area of controversy can be presented in terms intelligible to non-specialists. If research funding bodies were to insist upon peer review in appropriate cases before publication this would be one major step.

Educational research need not become devalued and manipulated for political or quasi-political ends, but the remedy lies ultimately within the research community. If that community fails to discipline itself, then the people who gain will be those with the best public relations, not those with the best research.