Research on Comprehensive Schools

Harvey Goldstein
The Professor of Statistics at the Institute of Education, University of London, previously worked at the National Children's Bureau and was a member of the DES advisory group which was given responsibility for advising the Secretary of State on the results of the comparative study of selective and non-selective secondary schools. Here he reviews this study, published last year under the title **Progress in Secondary Schools**, and discusses some criticisms of it.

Having more or less lost the political struggle against secondary comprehensive reorganisation, in the early 1970s the supporters of the selective system began to marshall a variety of statistical evidence in attempts to compare the achievements of children in selective and non-selective schools. Most of these attempts have used public examination results to compare achievements in grammar and comprehensive schools. One of the most sophisticated was that of Baldwin (Black Paper, 1977) who recognised that simple comparisons of comprehensive and grammar schools were invalid because the intake to many comprehensives had been 'creamed' by the grammar schools. Thus the better examination performance of grammar schools might be reflecting simply a higher ability intake. Baldwin, however, was unable to make a proper adjustment for intake differences because such measurements were unavailable, and instead he used an indirect method to estimate how the extent of creaming had changed over time and used this to 'adjust' his comparisons. Unfortunately, such a procedure fails to take account both of the increasingly selective nature of many grammar schools and of the fact that comprehensivation began first in Labour-controlled, and hence largely working class authorities, where one might expect in any case a lower performance.

At about this time, a new attempt to compare the performance of children in selective and non-selective schools was started. The Department of Education and Science, after hesitating for some time, funded a project at the National Children's Bureau using data collected in the National Child Development Study (NCDS). This study not only contained data on the educational attainments at 16 years of a large representative national sample of children but also had measures of attainment on the same children at 11 years, before transfer to secondary school. Thus the study was able not only to make comparisons between children attending different types of secondary schools, at the end of their compulsory schooling, it was also able to adjust for intake differences and thus obtain valid estimates of progress. Moreover, the children in the study were 11 in 1969 and 16 in 1974, a critical period during comprehensive reorganisation when sufficient schools of all types were in existence to provide useful comparisons. Thus, these data were the best available, or ever likely to be available, for comparing the different types of school. At the very least, therefore, even if the final results were to turn out equivocal, it should put a stop to any future attempts to use crude comparisons of examination results.

In July 1980 the results of the NCDS study were published, some of them having previously been leaked to the **Sunday Times** (16 March 1980). The report contains over 250 pages and is full of detailed statistical evidence and comment. I shall first summarise the main findings and then look at some of the political reactions to the report.

The principal results concerned tests of attainment in mathematics and reading at 11 and 16 years. For the purpose of comparisons between school types the analysis was restricted to those comprehensives which

had been comprehensive prior to 1969 when the children transferred to secondary school and likewise the grammar and secondary modern schools were those which had kept their status during the period 1969-1974. At 11 years, the scores of the children subsequently going to the different types of school predictably showed that those going to grammar schools had higher average test scores than those going to comprehensive schools who in turn had slightly higher average test scores than those going to secondary modern schools. Also, there was a marked social class difference, with about a quarter of comprehensive and secondary modern children coming from middle class (non-manual) backgrounds compared with just over half in the grammar schools. The crucial analysis was that which compared the average 16-year-old test scores in the three types of school, *at each level of eleven year attainment*. Thus the average 16 year reading test scores were compared for those with high 11 year reading scores, as well as for those with every other eleven year reading score. In this way, children with the same intake attainments at 11 in different types of school at 16 could be compared, constituting a valid allowance for the intake factor. In addition, social class and other variables were allowed for in this analysis. The main findings were as follows.

For both reading and mathematics, for those children in the 11 year old top 20% of the ability range, for given eleven year scores, the average scores of those in comprehensives and grammar schools were no different. While there were few secondary modern school children in this group, they did rather worse than those in grammar or comprehensive schools. Moving down to the next 20% of the 11-year-old ability range, the average score of the comprehensive children, for given eleven year score, comes closer to that of the secondary modern children with that of the grammar school children becoming rather higher than the rest. For the remaining 60% of the ability range the average score of the comprehensive school children is very nearly the same as that for the secondary modern school children at each 11 year score. The study carried out a number of other analyses on other aspects of schooling. It found that 16 year olds were more likely to stay away from school if they were in comprehensive or secondary moderns. The 16 year olds in secondary moderns had a higher reported rate of 'liking school' than those in comprehensives or grammar schools, whereas those in comprehensives were keener to stay on at school beyond 16 than those in secondary moderns with those in grammar schools the keenest of all. When questioned about choice of job, it seems that factors such as social class, parental education and sex were far more important than the type of school attended, which seemed to have little effect.

On reading the report one is struck by the care and thoroughness with which the issues are discussed. No attempt is made to oversimplify and the results are presented with all the qualifications they merit. It is pointed out that the study relates only to one historical and largely transitional period. Caveats are made about the particular measurements used, and there are reservations about the extent to which the study was able to make proper allowance for creaming. In my view, partly derived from being a member of the DES advisory group to the project, this research was as thorough and as competent as any in education. Most importantly, it demonstrated that *on the best available evidence*, there is little case for claiming that comprehensives are overall better or worse than selective schools or that they 'penalise' the high ability children. Of course, it must be admitted that no research is 'perfect'. Compromises usually have to be made and decisions have to be taken on the basis of accumulated experience and knowledge. While it is quite legitimate for critics to debate particular assumptions or conclusions, therefore, such a debate needs to be carried out responsibly, with a proper appreciation of the exigencies of research. In the remainder of this review I shall argue that much of the recent criticism of this research is in fact both misguided and irresponsible, and actually provides a useful object lesson in how not to carry on a public debate.

Two months after publication of the report, a long critique was produced by Cox and Marks (**Real Concern**, Centre for Policy Studies), and elaborated in the correspondence columns of the educational press. There seem to be four principal criticisms, focused on the analyses of mathematics and reading attainment. Firstly, they criticised the report for failing to set out the raw data. Secondly, they criticised the reading test in particular, on the grounds that it had too much of a 'ceiling' effect at 16 years, ie too many children were obtaining full or nearly full marks to make it a sensitive discriminator. Thirdly, they accused the author and research team of emphasising only those results which conformed to their predetermined view, thus presenting a biased and distorted report. Their final point, subsequently reiterated by Naylor (**Times Educational Supplement**, 22.11.1980), is that the grammar school children will tend to have had more test practice due to the presence of the 11+ which will have 'artificially' increased their scores, thus making their progress between 11 and 16 years look less than it really was.

On the question of presenting details of 'raw data' such as frequency distributions, there is an interesting area for debate. On the one hand, for experienced researchers such extra information should enable them to gain further insight into the results. On the other hand, report length is a consideration and it can be a fine judgement about what to include and leave out. Thus, although some disagreement is possible, any dispute is essentially marginal to the results themselves. Unless the critics argue that the omission per se actually threatens the validity of the conclusions (and Cox and Marks do not claim that), such a criticism is more in the nature of a debating point. To give it high prominence, especially in press correspondence, leaves the impression that they consider it to be a serious threat to validity. Such behaviour constitutes irresponsible criticism.

The issue of 'ceiling' efffects is somewhat more substantial than the previous one. The author of the report herself raises this problem but concludes that it does not invalidate the results. Nevertheless, such a view is a judgement and hence a potential source of weakness. Cox and Marks say about the use of the reading test at 16 that 'the National Children's Bureau knew the test was virtually useless at 16 for high attainers at 11 but presented the results in a way which conceals this fact'. Such a statement is not only immoderate and outside the normal canons of academic debate, it is also irresponsible, as before, because it gives the issue greater importance than it really deserves. In a response to Cox and marks (**Real Research**, by Jane

Steedman et al, National Children's Bureau), the full distribution of the reading test scores is given, from which it is clear that any 'ceiling' effect will not be serious.

According to Cox and Marks the report 'is so biased in its interpretation of its own data that it is hard to avoid the suspicion that those concerned with its production, including the Advisory Group on which the DES were represented, were capable of gross partiality and/or influenced by vested interests'. In subsequent letters to the press (eg **Times Educational Supplement**, 3.10.1980), this point was reiterated and emphasis given to the claim that the National Children's Bureau had made it difficult for them to gain access to the data, claiming that the research was 'shrouded in such secrecy and made so inaccessible' (**Guardian**, 7.10.1980). The issue here is not whether views about bias should or should not be entertained. Different researchers confronting the same analyses might well give different emphases which will reflect their own experience and views. To call this 'bias', however, with that word's popular overtones, is misleading. To come to different conclusions and to say so publicly does not exclude a constructive debate. In such a debate, however, the issues in dispute must be stated clearly so that, like all good criticism, the ordinary reader can understand and even take part in the controversy. Almost inevitably, once accusations are made such as the above, the real research issues tend to become lost and tempers rise. This then may alienate the interested bystander who will often declare a curse on both houses. A more responsible and moderate approach could easiy have been used which did not impugn the motives of the researcher. Furthermore, the data themselves were publicly available through the Social Science Research Council's Survey Archive so that it is incorrect to claim that they were inaccessible.

Finally, there is the question of the 'artificial' raising of the test score of the grammar school children. The implicit assumption behind this criticism is that, in some sense, those children whose scores are increased by practice will, if placed in a similar environment to children whose scores have not been so increased, show relatively less progress. This is, of course, a possibility, but there seems to be no real evidence to support it. Among other factors one would expect the quality of the later environment to be important. For example, if this is designed to enhance the skills which 11+ tests are meant to tap, which might be the case in many grammar schools, then one might expect the children exposed to practice to respond more favourably than the children who were not. The issue, therefore, remains an open one. Cox and Marks nowhere provide any discussion along these lines, but instead claim that the taking into account of their criticism 'would show the comprehensives in a less favourable light', and suggest that 'perhaps this is the reason why it has not been included'.

It is as yet too soon for many serious reviews of this research to have appeared in educational journals, although one fairly lengthy and balanced review by Colin Lacey (**Times Higher Education Supplement**, 16.1.81) attracted further responses from Naylor (**Times Higher Education Supplement**, 23.1.81) and Cox & Marks (**Times Higher Education Supplement**, 30.1.81) along similar lines to those referred to above. One minor academic criticism which had surfaced at the time of writing is a short article by Preece (in the **British Educational Research Association Newsletter**, December 1980). Although rather obscurely framed in terms of hypothetical IQ measurements, it raises the question of whether there are other intake measures which are relevant to the choice of school, for example, the number of older children in the child's family, which should have been allowed for in the analysis. The report itself does indeed recognise this problem and certain factors such as social class and sex were allowed for. It is to be hoped, nevertheless, that other researchers, who can now gain access to the data, will be able to pursue this topic.

I hope that, in defending this study against mistaken and irresponsible criticism, I have not given the impression that I believe it to be above criticism. It is no more immune from criticism than other competent pieces of research in education or social science more generally. As with other research, reanalysis by independent researchers is to be encouraged as one of the best safeguards against either incompetence or fraud. Until such work has been undertaken by the research community, however, it is my view that none of the criticisms so far made of the research poses any serious threat to its conclusions.