3 Studying Schools and their Effects

It has been shown that the proportion of school leavers who have attained a modest standard of basic number skills is considerably lower in Britain than in some other European countries, for example, West Germany.[1] This kind of comparison is rather limited. It is quite possible that there would not be the same contrast in the proportion of school leavers attaining a much higher level of skill, and that British schools, particularly those in the private sector, do as well or better at training an élite as those in other countries. However, a relatively low standard among the majority of children is a significant failing because of its implications for the personal development of the mass of the population and for the performance of the British economy. As an economic power, Britain has been in long-term secular decline since about 1870. There is wide agreement among economic historians that since industry has become dominant over agriculture, no country has transformed its economic performance without first radically improving its education system. In the present phase of development, fewer and fewer jobs requiring a minimum of education or training are available. Methods of working will continue to change rapidly, so people need to be equipped with the language, reasoning and number skills that will enable them to absorb new information and thereby adapt to change. These are strong arguments for the belief that making schools more effective should be a high priority for any present-day government.

However, until recently, the strongest tradition of thinking about school effectiveness has not been primarily concerned with improving personal development and economic performance. Instead, it has concentrated on inequality of attainment between individuals, and has examined the question whether better schools would significantly reduce these individual differences. This focus of thinking and research was a response to political programmes that saw education as a means of achieving greater equality; the two best examples of such programmes are the abolition of selection at 11+ and the Educational Priority Areas, an attempt to use extra educational resources to compensate for multiple deprivations in particular localities.

Two American studies strongly challenged the view that education could make an important difference to inequality: these were James Coleman's report, published in 1966, on *Equality of Educational Opportunity*[2] and Christopher Jencks's reanalysis of Coleman's and other material, published in 1972 in the book *Inequality: a reassessment of the effect of family and schooling in America*. Jencks concluded that 'equalizing the quality of high schools would reduce cognitive inequality by one per cent or less' and that 'additional school expenditures are unlikely to increase achievements, and redistributing resources

will not reduce test score inequality'.[3] Essentially, the basis of this claim was that a child's test scores or examination results could be predicted far more accurately from knowing the family background than from knowing which school the child went to. From the way the conclusion was phrased, it is clear that it was a direct response to the compensatory education programmes that were favoured in the 1960s both in the US and in Britain. However, it was interpreted as meaning that education had no effect. As Michael Rutter has put it 'these conclusions were both widely quoted and interpreted as meaning that schooling had such minor marginal effects that the educational process was scarcely worth the relatively large resources poured into it'.[4] Whereas Coleman and then Jencks had concluded that schooling was not effective as a means of reducing individual inequality, their conclusions were taken to mean that schooling has little influence on whether or not children can read, write and do arithmetic.

Over the past ten years, this view has increasingly been challenged, both in America and in Britain. Our own research is one of a number of projects that have set out to demonstrate that different schools have different effects, and have looked for explanations of these differences. The two most important British projects are *Fifteen Thousand Hours*, by Michael Rutter and his colleagues, a study of 12 secondary schools published in 1979; and *School Matters*, by Peter Mortimore and his colleagues, a study of 50 junior schools published in 1988. There is, in addition, a large body of other research and writing that is concerned with effective schooling and related matters. In 1983, Purkey and Smith produced a review of the American evidence, and Michael Rutter produced a review of American and British evidence.[5] Here we will not try to summarise the results of past research. Instead, we will make some general points about the analysis of school effectiveness, and about the associated methods of research. It would be good to be able to say that this discussion of the problems involved in understanding school effectiveness prepares the way for the presentation of findings, which will show how these problems can be solved. However, that would be a distortion on two counts. First, our research was only partially successful in terms of the objectives as originally defined, because we met with serious practical difficulties which could not all be overcome within the time and resources available. More important, to the extent that a conceptual framework can now be sketched, this is an outcome of the project. It is not the same as the framework that shaped the original design.

The meaning of effects

The basic problem in measuring the effectiveness of schools is that the level of attainment reached by children is influenced by a wide range of factors in addition to the school. Hence it cannot be assumed that the attainment of the children (or their responsiveness or sense of social responsibility) is a product of their schooling, or, therefore, a measure of its effects.

This problem is common to all attempts to measure the effectiveness of social institutions. In fact, in most other cases, the links between the activities of the institutions and the desired results are far more difficult to trace. For example, the police aim to keep the peace, to prevent crime and to catch and prosecute offenders. However, the level of public order, the crime rate and the proportion of offenders who are caught are all the result of complex social processes. Among all these influences, it is extremely difficult to trace the effects of policing.

In looking for a way of defining the effects of institutions, one approach is to consider how far the outcomes would be changed if the institution did not exist at all. At first, this

sounds like a definition of the total effect of an institution. However, it is not very useful in practical terms. According to Rutter, the available evidence suggests that children's cognitive development is, in fact, affected by whether or not they attend school. Also, 'the evidence on the effects of school closure in Western societies, either as a result of war... or attempts to avoid racial integration... are in keeping with that conclusion - as are the findings on the benefits of continued schooling during late adolescence'.[6] But it is difficult to use this kind of evidence in an ordered way to provide a quantitative assessment of the effects of normal schooling; and of course this approach is no use at all in showing what kinds of schooling work best, and why. Apart from the practical problems in obtaining good evidence, there is an interesting structural problem with this line of argument. A society minus a particular institution is one of two things: either it is managing temporarily without that institution (as in the case of school closures to avoid racial integration) or it is an entirely different kind of society that does not value what that institution does, or achieves it by entirely different means. An example of the second case would be tribal societies without writing or mathematics which pass on an extensive repertoire of skills, knowledge and oral history and secure close cooperation from children without any formal schooling. In the case where schools are temporarily closed, they may be partially replaced by something else; for example, parents may teach their children at home. More generally, other social processes and institutions will not remain fixed when the schools are temporarily closed, but will adapt to the new situation. In the case of a society that has not yet developed schools, it will be radically different in so many ways from a more technically developed and socially differentiated society that it would be difficult to make any useful comparison. It may seem to be common sense to think of the effect of schools as the difference it would make if there were none, but on closer analysis it is hard to attach any definite meaning to this idea, either in principle or in practice.

Most research and analysis of effectiveness adopts one of the many approaches having an important family resemblance: that they try to measure or analyse the differences between the results achieved by different schools (or other institutions). How 'the results achieved by different schools' is defined and analysed is crucial (the point will be discussed below). But given an appropriate definition and analysis, studying the differences between the results achieved by different schools is a promising approach, and it does bear on the most general questions of effectiveness. If no difference could be found between the results achieved even by schools adopting contrasting styles of teaching and methods of organisation, then this would, indeed, suggest that schools have little or no effect. The reason is that if schooling does have some effect, then the amount and nature of the effect is bound to vary between one school and another, unless the schools are extraordinarily homogeneous. This is enough to show that school differences are a promising starting point for the analysis of school effectiveness. But there remains a problem in defining how much effect a school has.

The persistence of mental capacities

The groundwork for research on school effectiveness has been done by the mental testing movement. A huge range of test instruments has been developed to measure many different mental skills and aptitudes. Compared with the measures used in much social research, the best of these instruments have been developed by rigorous and refined methods. They have been shown to have a high level of reliability (the same person, if tested twice, would obtain closely similar scores). There is also a huge body of information relating to the validity of the main tests, which have been used in many different studies; for example,

different tests of reading comprehension do appear to be measuring reading comprehension, because the score that a person obtains on one such test is highly correlated with the score he obtains on another.

Discussion of the subject of mental testing has been clouded by the controversy over 'I.Q. tests' and 'intelligence'. I.Q. tests are merely composite tests that cover a range of mental skills and depend as little as possible on knowledge of any specific subject matter. There is no particular reason to describe them as tests of intelligence, except that scores on the component sub-tests, which cover more specific mental skills such as verbal or non-verbal reasoning, are quite highly correlated, so that, for example, someone who has a high verbal reasoning score tends quite strongly to have a high non-verbal reasoning score too.

I.Q. tests have been discredited among many sociologists because they were associated with the theory that intelligence (or effectively the I.Q. test score) is largely determined by genetic factors. For the purpose of analysing school effectiveness, there is no need to enter into this controversy. It is important to recognise, however, that independently of their theory of the origin of individual differences, the mental testers had discovered at least two very important facts: that mental capacities vary widely between individuals, and that there is a very strong tendency for the mental capacities of an individual adult to persist over time and for the mental capacities of an individual child to develop in a predictable way. With the use of the refined and reliable instruments that have been developed, if a child is tested in mathematics and reading at the age of eight, it is possible to make a surprisingly good prediction of the scores that the same child will achieve in mathematics and English (but on different tests) at the age of 15. Furthermore, in seeking to predict the child's test scores at 15, a knowledge of his or her scores at eight will be far more useful than any other piece of information; much more useful, for example, than knowing what socio-economic group the family belongs to or, indeed, what kind of school the child attends.

Thus mental capacities develop in a relatively predictable way, and having developed they tend strongly to persist. Also, individual differences in mental capacities are strongly related to family background (how far this is because of genetic inheritance is not important for the present discussion). Because children's test scores tend to grow in a predictable way, a test score at any one age (from about five onwards) is a very good basis for predicting the score at a later age. Family background factors are a fairly good basis for predicting the scores obtained on a range of tests at any one age. At the same time, of course, a substantial part of the variation in scores between individual children is not related to family background: there are many low-scoring children coming from privileged families, and many high-scoring children coming from disadvantaged families. Still, family background is quite a powerful predictor.

Jencks et al. concluded that about 50 per cent of the variance between individual children in scholastic attainment was attributable to family background, while only 2 to 3 per cent was attributable to school variables. Essentially, this was a re-statement of the very well-known facts summarised above. Individual differences in attainment are large; attainment is strongly related to family background; and individual performance shows considerable stability throughout the process of child development. School effects will appear small in relation to the very wide range of highly stable individual differences.[7]

It is correct to conclude from this (as Jencks did) that improving the standard of schooling in the poorer schools will not significantly reduce individual inequalities. (As Rutter points out, lowering the standards of the better schools would probably be a rather better method of reducing individual differences, but even so it would not reduce them

much.) However, it does not in the least follow from this that schools have little effect, or that an improvement in school effectiveness (by an amount equivalent to the difference between the best and the poorest schools) would not produce a significant improvement in attainment. There are two reasons for this.

1. If there were no school differences this would suggest that schools have little effect, but it is not true that the proportion of variance attributable to schools is a good indicator of the size of school effects. This is because the proportion of variance attributable to schools depends on how heterogeneous the schools are. There are strong pressures towards homogeneity (such as school inspectors) and it is likely that many schools have much the same effects, but a few schools are much better or much worse than the majority. The difference between the best and worst schools would be a much better indicator of the effects of schooling than the proportion of variance attributable to schools, since this latter figure would be mostly determined by the mass of relatively homogeneous middling schools. In any case, no measure of school differences is a quantitative measure of the total effect of schooling, since the schools actually existing at any time do not represent the complete range from the best possible schooling to no effective schooling at all.
2. There is no justification for using individual differences as the yardstick for measuring school differences. Differences between schools may be small compared with differences between individuals, but may still be very important. A small increment in the average score in mental arithmetic, for example, may take a large proportion of individuals across the threshold needed to be able to retain a score at darts.

A new approach to modelling school effectiveness

Our research project is one of three that adopt a new approach to modelling school effectiveness (the other two are the ILEA junior school project, published as *School Matters*, and the Thomas Coram infant school project). This new approach developed from the pioneering work of *Fifteen Thousand Hours*, but incorporates important improvements in statistical procedures developed through the work of Murray Aitkin and Nicholas Longford, formerly of the University of Lancaster, and Harvey Goldstein of the University of London.[8] The method depends on having longitudinal data about a group of children in a number of schools. There has to be a measure of attainment (say a reading test score) for each child at one time, then another measure (say another reading test score, or a set of examination results) for each child at a later time. There must be information about the family background of each child. A mathematical model is set up to predict the later measure of attainment, for each individual child, from the combined information about the earlier measure of attainment and family background factors. The model assumes that the later measure of attainment may vary according to the school the child belongs to, and that the various relationships (for example, between early and later attainment) may vary between schools.

Because attainment at an earlier time is one of the factors used to predict later attainment, this procedure is roughly equivalent to analysing the differences between schools in the *progress* achieved by children with similar initial attainment and other characteristics. It turns out that family background factors explain a relatively small proportion of the variance in *progress* between individual children (whereas, as Jencks pointed out, they explain a relatively large proportion of the variance in attainment at any one time). Within the terms of this kind of model, school differences are at least as important as family

background factors, but earlier attainment is much more important than either. Thus, if we discount the very strong tendency for the level of attainment to be stable over time for the same child, school differences are seen to be highly important compared to other factors. This is another way of saying that children with a given level of initial attainment achieve markedly different rates of progress depending on which school they go to. It remains true that these school differences are small compared to the large differences in initial (and final) attainment between individual children. Nevertheless, these differences are large in terms of the standard of education that the schools are delivering.

A feature of these procedures is that they allow for regression analysis (as described above) with multi-level data. In the case of the present project, there are just two levels: the child and the school. In the ILEA junior school project, there are three levels: the child, the class and the school. Information at the school level can be used in the analysis, as well as information at the child level. This means that we can show how much of the difference in progress between schools is associated with some characteristic of the school.

For the first time, these methods provide a framework for analysing school effectiveness in a rigorous way. The methods are much more advanced than those available for studying the effectiveness of most other kinds of organisation. This is because some, at least, of the objectives of schooling are very well defined, and because of the achievements of the mental testing movement, which has provided some of the necessary instruments. Nevertheless, there are still substantial difficulties in trying to understand the processes underlying school effectiveness. The rest of this chapter is an attempt to explain why.

Research design and the size of school effects

The method just described allows us to make valid comparisons between the outcomes achieved by schools having different intakes, in terms of attainment, family background and so on. Another approach to the problem of making valid comparisons is to find two schools that are 'matched' or closely similar in terms of the attainment and family background of the children entering them, then compare the test scores or examination results of children in these schools at a later age. This method was used in a number of earlier American studies. There are great difficulties in finding matched pairs of schools, and it is not possible to match on more than a couple of characteristics, so the intakes of any 'matched' pair of schools will usually differ significantly with respect to characteristics that were not brought into the matching process. In principle, multi-level multiple regression procedures allow us to overcome these problems, by providing a method of validly comparing unmatched schools.

Nevertheless, there remain some problems of research design. It is important to recognise that the apparent size of the school effects will depend on the way the sample of schools is selected. If they are relatively homogeneous, or falling within a restricted range, then this will tend to reduce the size of the observed school effects. In practice, none of the main studies has covered the full range of schools. *Fifteen Thousand Hours* covered 12 local authority comprehensives in an inner city area, probably the most homogeneous sample of the four main studies. *School Matters* covered a representative sample of ILEA primary schools. This sample is considerably more inclusive than the one in *Fifteen Thousand Hours*, first because it covers the whole of inner London (as opposed to one borough) and second because it covers the full range of local authority primary schools in inner London (whereas there were selective secondary schools in the public sector not covered by *Fifteen Thousand Hours*). The Thomas Coram project covered 33 infant schools in multi-racial areas of inner London. As explained in the next chapter, in our own

study we deliberately set out to select a heterogeneous set of comprehensives, and we covered four local education authorities, so of these four studies, ours was probably the one with the most heterogeneous sample of schools. At the same time, all four studies exclude the private sector of education, and none of them covers selective schools of any kind.

We have already argued that the size of school differences is not a direct indicator of the total effect of schooling. It should further be recognised that the existing studies cover only a restricted range of schools. It would probably be possible to demonstrate larger differences in school effects if a single study could encompass schools of all types.

However, there are penalties as well as advantages connected with having a heterogeneous sample of schools. The more unalike the schools are, the more of a strain this puts on the statistical procedures used to compare like with like. Some of the schools will have unusual characteristics, or combinations of characteristics, which make it difficult to generalise from their results. They are all different, but not in a systematic way, so various possible types are not represented at all (for example, in our own study, there is no school with a high proportion of Sikhs, and no school with a high proportion of Bangladeshi boys).

These are the limitations of general purpose, descriptive studies, which aim to cover a fairly broad range of schools, and do not explicitly set out to test a theory or explanation of school differences. They are something of an uneasy compromise. They do not measure the full extent of school differences (because of the important exclusion of selective schools and the private sector). They are intended to be capable of testing many sorts of explanations of school differences, but there are many particular explanations that they cannot test, because this would require the selection of schools representing particular types which would serve as a test of the theory.

Testing theories

Both *Fifteen Thousand Hours* and *School Matters* are based on studies that collected an impressive array of information about schools and what happened in them. In *School Matters*, both school level and class level variables were examined for their relationships with school effectiveness. 'Given' variables at the school level were the building, resources, intake and stability of staff and pupils. 'Policy' variables at the school level were the Head's style of leadership, the type of organization, the involvement of staff, the curriculum, the rewards and punishments used, parental involvement with the school, equal opportunities, the school atmosphere. 'Given' variables at the level of the class were number of pupils, the age, social class and ability composition of the class, the classroom, resources, curriculum guidelines, teacher characteristics, whether the teacher changed during the year. Finally, the 'policy' variables at the class level were the aims and planning of the teacher, the teacher's strategies and organization of the curriculum, management of the classroom, including rewards and punishments, classroom atmosphere, the level and type of communication between teacher and pupils, parental involvement, and record keeping.

It should be mentioned, in passing, that most of these 'variables' are really complex structures or policies or processes that have been summarised in some way by the researchers. The analysis is in three parts. First, the authors describe the pattern of relationships between each of the school and class level variables and the measures of school effectiveness that have already been established. There are very many relationships of this kind. Next, the authors discuss the pattern of relationships between the school and class level variables themselves. This suggests theories about what is going on (say about

how head teachers influence the way teachers plan their lessons) but the whole pattern of relationships is highly complex; many different theories could be put forward to explain it. Finally, the authors identify 'twelve key factors' that lead to an effective school. They say that this identification of twelve key factors is informed by the foregoing analysis, but not determined by it.

The main point to be made about this analysis is that it is alarming rather than helpful to find such a large number of relationships, unless there is a way of understanding how they all fit together. Without a theory of how schools work, contemplating this immensely complex pattern produces not enlightenment but a cognitive snowstorm. The authors of *School Matters* recognise this problem. When they come to identify the 'twelve key factors' they acknowledge that 'these factors are not purely statistical constructs. They have not been obtained purely by means of quantitative analyses. Rather, they are derived from a combination of careful examination and discussion of the statistical findings, and the use of educational and research judgement. They represent the interpretation of the research results by an inter-disciplinary team of researchers and teachers' (p248). This is equivalent to saying that they have used the results of the research to help them formulate a theory about how schools work (and what makes them work effectively), but that those results do not act as a test of the theory they have formulated.

One of the main reasons for this is that each of the school and class variables is related to many of the others, but at the same time to the measures of effectiveness. This means that it is often not clear which aspects of school policies and practices are the critical ones. The most interesting conclusions arise when there is no relationship between some school or class variable and effectiveness. For example, it was found that 'the amount of teacher time spent interacting with the class (rather than with individuals or groups) had a significant positive relationship with progress... In contrast, where a very high proportion of the teachers' time was spent communicating with individual pupils, a negative impact was recorded... Measures of the extent to which a whole class teaching approach was adopted were very weakly and not significantly related to progress... It was the number of interactions *involving* the whole class, rather than any attempt to teach the whole class as one unit, that seems to be have been associated with beneficial effects.'(p228) Because some of the variables here were negatively or not significantly related with progress, the findings tend to support a specific theory (namely that the total amount of useful contact between teachers and pupils is the important factor). However, in many other cases, a whole set of policies are related to each other and individually to progress, so it is hard to know which ones are critical.

Certainly there are theories about how schools work and what makes them effective, which are associated with the recent research. Michael Rutter puts forward one general theory and a number of specific ones in his review article published in 1983. The general theory is that while individual teachers vary in their effectiveness, and while effects depend partly on the details of the curriculum, there is something called the 'school ethos', a set of schoolwide influences that make it more or less likely that teachers will conduct their lessons in an effective manner. An example of one of Rutter's more specific theories is that the way teachers manage the classroom is crucial to effectiveness; this he sees as essentially a matter of maximising the amount of time the pupils are engaging in useful learning or practice, and the teacher does this, for example, by engaging their attention, securing orderly behaviour and managing his or her own behaviour so as to maximise the amount of useful contact with each member of the class. He also points out that this theory

has important policy implications, since teachers are given little instruction or training in classroom management, and are not helped to develop these skills on the job.

Evidence can, of course, be cited in support of both the general and the more specific theory mentioned above.[9] Nevertheless, these theories have certainly not been tested exhaustively, and in fact research has not been designed to test them in a rigorous manner, but rather to describe a vast web of relationships. Excellent techniques of research are now available, and some theories about school effectiveness have been quite clearly articulated. In the future, research should concentrate on testing clearly articulated theories in a more focused way.

Static versus dynamic conceptions of the school

The analytic model used for schools effectiveness research tends to be associated with a static conception of the school. Although the information about the schools is obtained over a considerable period of time, each measure is presented as though it described the school at one instant, and there is no attempt to describe any process of change. Instead, the apparently timeless characteristics of effective schools are compared and contrasted with those of less effective ones.

The limitations of this approach are most obvious when methods of management are under consideration. Both *Fifteen Thousand Hours* and *School Matters* show strong relationships between the style of management adopted by the head teacher and school effectiveness. This raises two related questions. One is about the direction of causation. It is not clear whether the head is able to adopt a given style of management (for example, ask teachers to provide records of children's work) because teachers are competent, relationships are good, and such requests can readily be met; or whether the standard of teaching and quality of relationships has been improved because the head has requested records of children's work. The second kind of explanation - that good management at the top makes all the difference - is of course irresistible to top managers like the heads of university departments who direct this kind of research: but it may be wrong.

The second related point is that schools, like other organisations, go through periods of relative stability followed by shortish periods of sudden change. The research will catch most schools in a period of relative stability. Let us suppose that the style of management by the head is, in fact, crucial in shaping the school. If the school is functioning badly, there will be a style of management required to bring about a series of rapid changes, transforming the school into a good one. If the school is functioning well, there will be a style of management required to maintain stability. These two styles may well be entirely different. What research has tended to observe is the management styles associated with stable effective states. This says little about the styles required to transform a bad school into a good one, which would probably be entirely different. Assuming that what the head does is, in fact, critical in transforming a bad school, it is still quite possible that the head's style has little importance in maintaining an already good school, and that the causation in that case runs in the opposite direction.

This point can be summarised by making a distinction between the conditions existing in an effective school, and the actions that have to be taken to transform an ineffective school into an effective one, or to maintain the performance of an already effective one. These two sets of actions may not be the same.

Clearly, theories of how schools work need to be developed to take account of change and the maintenance of a stable state. These considerations suggest the need for a different kind of research, that studies schools as they change, and perhaps research that observes

the results of taking specific actions. However, even within the framework of the school differences model here discussed, there may be some scope for testing explanations that take in organisational dynamics.

Notes

1. See Prais and Wagner (1985).
2. See Coleman et al. (1966).
3. See Jencks et al. (1972), p109.
4. See Rutter (1983), p1.
5. See Purkey and Smith (1983); Rutter (1983).
6. Rutter (1983), p2.
7. For a variety of reasons, however, Jencks et al. produced a very low estimate of the effects of school differences, even within the terms of their own argument. For example, they considered the difference in results achieved by different *categories* of schools, classified in various ways (for example, according to expenditure). The differences between the results achieved by individual schools might have been much greater; and the classification of the schools was not based on an understanding of the factors that actually influence effectiveness. It would not be surprising to find little difference between the categories of schools if the effective ones were fairly evenly divided between these categories.
8. See Aitkin and Longford (1986), and Goldstein (1987).
9. There is much evidence that the amounts of time spent in class and in useful learning ('on task') vary substantially between schools, and some evidence that they are related to achievement. For a review of the relevant research, see Bennett (1978).