

PART III SCHOOL EFFECTS AT THE END OF THE SECOND YEAR

9 Attainment in Absolute Terms

In Chapter 3 we have discussed the substantial difficulties that arise in trying to assess the effect of schooling. The general approach used in this study is to measure the differences between the effects achieved by different schools. We have seen in Chapter 5 that there are large differences between the pool of pupils entering the study schools, even though these schools were chosen from within a restricted range. For that reason, the absolute level of attainment of children at later stages of their careers is not a valid measure of the success of the different schools. Instead, in the next and subsequent chapters, we shall describe the use of a mathematical model which predicts attainment at a later time, for each individual child, from the combined information about attainment at an earlier time and family background factors. This procedure is roughly equivalent to analyzing the differences between schools in the *progress* achieved by children with similar initial attainment and other characteristics.

However, when setting up any multivariate model, it is necessary to make many simplifying assumptions. Before starting to simplify, it is important to study the results in a more flexible and open-minded way. That study will inform decisions about the multivariate analysis and suggest necessary qualifications to the results it produces. The present chapter therefore considers the results of the attainment tests in absolute terms, and the pattern of relationships between attainment and a range of background factors.

The tests

First year

1. NFER reading comprehension test, maximum score 15. This test was of the kind where the child reads a short passage then answers questions about it. Three passages were included as against the usual four (because of the limited classroom time available for this testing). The questions were intended to test propositional content (extraction of information from the passage), extra-textual knowledge (information brought to bear on the passage from outside the text) and style (understanding of the purpose of the

passage). However, the test does not produce separate scores for different skills. The questions are all in the same format, but a variety of skills are needed to answer any one of them.

2. NFER writing test, marked on a score of 1-7. This was a specially commissioned test of 'free writing' taking themes from the passages in the reading test as the starting point. The method of assessment used by NFER was described as 'rapid impression marking'. It is not possible to say exactly what is being assessed, but certainly not simply spelling and grammar.
3. NFER maths test of 30 items. This was a specially commissioned test made up of items drawn from the 'LEA and schools item bank'. Maths advisers from the four study areas were consulted about the balance of items as between different maths subjects, in the light of the emphasis given to the various subjects in primary schools.
4. NFER non-verbal reasoning test of 25 items.
5. NFER verbal reasoning test of 25 items.

The two reasoning tests were specially commissioned and made up from items drawn from the NFER 'LEA and schools item bank'.

The maths and reasoning tests were shorter than the standard tests used by NFER, so as to fit the amount of classroom time available.

Second year

1. The Edinburgh Reading Test Stage 4 was used, which is designed for children aged 12 to 16. This is a much more detailed and varied test than the NFER reading comprehension tests or indeed any others that are available. It requires two sessions of 30 minutes. The maximum overall score is 155, but within this there are five sub-tests that rely on different types of material and question.
 - a) Skimming(scores 30)
 - b) Vocabulary(scores 35)
 - c) Reading for facts(scores 30)
 - d) Points of view(scores 35)
 - e) Comprehension(scores 25)

2. NFER maths attainment test EF, maximum score 60. This is a standard test for children aged 11 to 13:06. Although it is not strictly timed, it requires a maximum of 50 minutes.
3. NFER verbal reasoning test EF, maximum score 90. This again is a standard test for children aged 11 to 13:06. It takes 40 minutes.

Discussion of testing methods

The tests used in the first year were the subject of considerable criticism at a number of the schools. Some of the most important points made were that they were too difficult, that they were culturally biased, that they were discouraging or even disturbing to less able children, and that they were not reflecting the whole educational experience, in all its richness and diversity, that the schools were aiming to impart. All of these criticisms had some truth in them, and it is worth discussing each one in turn.

Level of difficulty

The reading and verbal reasoning tests used in the first year were too difficult for the children at the study schools, in that the overall means were lower than expected (see Table 9.1), an appreciable number of children scored zero and a considerable proportion achieved only very low scores. From the technical point of view, the loss here is a lack of discrimination at the low end of the range. From the human point of view, it means that an appreciable number of children could not really attempt the tests at all, or could only answer one or two questions, and undoubtedly this was discouraging and even distressing in a number of cases. The problem arises partly because NFER tend to standardise their tests in relatively 'easy' schools or areas which contain a relatively low and certainly not an above average proportion of children who have problems or find tests difficult; while, by contrast, the study schools tend to contain an above average number of children with difficulties, some of them linguistic or otherwise connected with adaptation to British culture. But to some extent this research problem is probably insurmountable, since no single test could adequately discriminate across the whole range of performance; if the tests were made easier there would be a lack of discrimination at the top end of the range. Because the tests used in the second year contained more items,

however, the problem of discrimination at the lower end of the range was partly overcome, although there were still some zero scores.

Cultural bias

It is important to analyze what is meant by 'cultural bias'. Any test (indeed, any text, discourse or act of understanding) exists within the set of assumptions and responses that constitute a given culture, so that the idea of a 'culture-free' test is, truly, nonsense. This means that members of a minority group cannot be tested in a common frame with members of the majority unless they have some of the assumptions and responses that belong to the majority culture, and how well they do is bound to be partly a reflection of how well they have learnt to work within the majority culture. It seems important that this point should be firmly established, otherwise people will expect and demand a yardstick of progress or achievement that is independent of the content or values of a particular culture, and such a demand can never be met.

But this is to consider the matter very broadly for the sake of making a general point. Closer consideration will immediately show that the 'culture' or 'majority culture' is an all-embracing concept: so much so that the 'majority' are ignorant of much of the 'majority culture', which from another viewpoint consists of a vast number of partly overlapping sub-cultures shared by people with some common interest, experience or way of life. It therefore makes sense to distinguish between those elements of our culture that we consider to be 'central' in which we expect everyone to be able to share, and those which belong to a particular social or interest group. The central part is the common currency of cultural exchange, the part that people need, at a minimum, to be able to support themselves and relate to others in most social contexts. Historically, this central element of the common culture may derive from the dominant group more than from subservient ones – the latest form of the common language may be most like the language of the latest wave of invaders – yet what makes it central is not that it belongs to the dominant group but that it is held in common by most groups, that it has become a *lingua franca*.

It follows that although a test cannot be 'culture-free' it may be 'culturally biased', if that is taken to mean that it refers to or relies on aspects of culture that are incidental or of specialist interest or belong

to a sub-culture and not to the central core. Of course, there is room for many disagreements in detail about what is 'central', but we would accept that in a few cases the first-year tests made specific references that would better have been avoided and which did not belong to the central culture: for example, one of the verbal reasoning items referred to 'fairy rings'. However, references of this kind in the tests were very few, too few to be significant in the context of the tests as a whole. An item analysis of the first-year reasoning and reading tests was carried out on a sub-sample of the scripts, taking into account differences between ethnic groups (classified from information given by some of the schools). This showed that for a number of items on the verbal reasoning and reading tests there were substantial differences between ethnic groups in the proportion getting the answer right, even though most of these items did not rely on cultural references that were obviously outside the 'common core'. Differences of this kind were much less in the case of the non-verbal reasoning test, which suggests that the cultural differences that are significant in this context are mostly connected with language.

We conclude that, in the sense in which we have used the term here, there was a degree of cultural bias in the first-year tests, though this was probably not very significant. At the same time, there were large differences in performance between ethnic groups, especially on the verbal reasoning and English tests and on certain items. This is, of course, connected with cultural differences, and especially with the fact that a substantial proportion of the children in certain groups had limited English at the time, but not with a cultural bias in the tests themselves.

Scope of the tests

The final criticism of the first-year tests that we have mentioned is that they were not reflecting the whole educational experience that schools are aiming to impart. This is, of course, entirely true. It seemed best to concentrate on the most basic elements of education which are most likely to be held in common by different schools. In this context it is extremely important that the schools themselves show, by their actions, that the basic skills are at the core of what they are trying to achieve, and that they believe they can be tested. Almost 80 per cent of local education authorities currently operate some kind of standardised

testing programme, and reading is the skill most commonly tested. Indeed, the whole 'testing industry' has grown up in response to the demand from the education system, which uses tests, of precisely the kind that were used in this study, on a very large scale. Thus, although tests only touch a part of what schools are trying to achieve, it is clear that this is a part that schools and education authorities consider to be of central importance.

Change of emphasis in the second-year tests

A number of changes were made in the tests used in the second year compared with those used in the first. The admittedly experimental test of free writing was dropped, largely because the marking criteria seemed too subjective and could not be made very explicit. We felt that it was particularly important to have a more detailed, varied and widely-based test of reading, and for that reason changed to the Edinburgh Reading Test. The test of non-verbal reasoning was dropped for the second year because the measure obtained in the first year seemed satisfactory, and this was not a skill that would be expected to respond in a direct way to teaching or other aspects of the school environment; in other words, the non-verbal reasoning score is more useful as a 'control' variable than as a possible measure of progress. Because there were doubts about the verbal reasoning test used in the first year, a longer and more detailed test was included in the second year. The maths test used in the second year was essentially similar to the first-year one, but was longer and more detailed.

The general plan was that academic performance would be measured in terms of the maths and reading scores, with the reasoning scores acting as useful 'control' variables. The tests used at the two points are not entirely comparable, especially in the case of the reading tests. This is not of critical importance, since the plan does not rely on a concept of 'progress' on given scales; instead, it relies on discounting differences in the second-year scores that could have been predicted from the differences in the first-year scores.

Findings

The overall means and standard deviations for each of the eight tests, together with the numbers of children tested, are shown in Table 9.1.

These statistics show that in the case of the first-year reading and verbal reasoning tests there is poor discrimination at the low end of the range, while the distributions for the other three first-year tests are much more satisfactory. The statistics suggest that the second-year maths and reading tests provide good discrimination across the whole range, but the verbal reasoning test again provides poor discrimination at the lower end, even though the study children were towards the upper end of the age range for which it was designed (11 to 13:06). Nevertheless, the second-year verbal reasoning test provides a more sensitive measure than the shorter test used in the first year, because of the larger number of items.

For ease of making comparisons between tests and between the first and second years, the scores have been standardised *on the population of children tested in this study*. The standard score, or 'z score', expresses the score for any individual or group as the deviation from the overall mean in units of standard deviation. It is helpful to bear in mind, when interpreting these standard scores, that roughly 95 per cent of the individual scores will lie within a range of 2 standard deviations on either side of the mean, while about two-thirds will lie within a range of 1 standard deviation each way. This of course implies that a difference of, say, 0.5 between two standardised scores is a very large difference. It is also useful to bear in mind that because the reference point is the overall mean (for the study children) any group that has a standardised score that is significantly different from zero is showing a different result from the study children as a whole. It is important to emphasise that in this case standardisation is purely a device to make comparable the scores from the different tests within the study: the standardised scores carry no implication about how the children within the study compare with a wider population.

Analysis by school (Tables 9.2 and 9.6)

Large differences are shown between the mean scores of children attending different schools, both at the beginning of first year and at the end of the second year. This is not surprising in view of the large differences between the schools in the ethnic and social class composition of the children. In general, the scores are lowest within

area 1 and nearly as low within area 4, while they are highest within area 2.

Looking at the scores for the individual schools, we are faced with a very complex picture. However, a few points can usefully be made at this stage. The first-year scores seem to show that the correlation between reasoning and attainment scores is far from perfect. At some schools (for example, school 12) children do much worse on reading, writing and maths than on non-verbal or verbal reasoning, while at others (for example, several in area 2) the opposite is, if anything, the case.

The mean scores on reading comprehension were very low in the first year at a number of schools (15, 12, 42, 14). The school with the lowest first-year reading score (15) is the one with the highest proportion of bilingual children and of children who came to Britain since 1976, while the schools with the second and third lowest first-year reading scores contain middling proportions of bilingual children. However, the low reading scores among the intake to some schools are only partly related to the presence of children who are struggling to learn English as a second language.

The girls at school 15 clearly stand out as having particular problems, probably connected with lack of English and recency of arrival in Britain. However, it appears that there was no relative improvement in their performance over the two years: in fact, their score in reading got distinctly worse, and their maths score remained extremely low. The fact that the verbal reasoning score at this school went up suggests that the problems may be specifically related to lack of English.

Comparison between the first and second-year reading and maths scores shows some potentially interesting differences between the schools. It seems particularly significant that changes in the maths score are often quite different from changes in the reading score, which implies that some schools make much better progress in the one subject than in the other.

Analysis by country of origin, sex and religion (Tables 9.3 and 9.7)

Looking at the first-year results, we find that Bangladeshis had much lower scores than any other group; perhaps surprisingly, their lowest

score was in maths (-1.120) rather than in reading (-0.705). Even these figures under-state the low level of initial performance of Bangladeshi children, since an appreciable number of those in school 15 could not attempt the reading test and are therefore excluded from the tables. Pakistanis were the group with the next lowest scores (reading -0.489, maths -0.501). Indians and 'other' south Asians did better in the first year, but were also below average. Of all the south Asians, African Asians do best, achieving first-year scores just below the average. Children of West Indian origin scored below average in the first year, although the difference from the mean is not enormous (reading -0.21, maths -0.28). Thus, all of the groups originating from outside Britain have below-average scores, so that children of British origin must have above-average scores.

West Indian girls scored distinctly higher than West Indian boys on all five first-year tests: in fact, the West Indian girls were only slightly below average, while the boys were distinctly below average. There are other differences, too, between the first year scores achieved by girls and boys, but these do not much affect the broad pattern of differences between the ethnic groups.

At the end of the second year, this pattern remained roughly the same, although the Bangladeshis appear to have fallen further behind in reading (possibly because Bangladeshis who could not read English at all were not tested at the beginning of the first year). The gap between West Indian boys and girls had widened, and the boys are scoring well below average, though not as low as Bangladeshis. West Indian girls, however, were scoring around the average at the end of the second year, but lower than white British girls, who were scoring slightly above average at that point.

The analysis by religion cuts across the one by country of origin. All of the Pakistanis and Bangladeshis are Moslem, most of the African Asians are Hindu, while the Indians may be Moslem, Hindu or Sikh. There are also Moslems originating from countries outside the Indian sub-continent. The Moslems scored considerably lower than other religious groups in both years. The Hindu and Sikhs scored around the average, but the Hindu slightly higher than the Sikhs. It is of some interest that the small number of Jews in the sample achieved remarkably high scores.

South Asians: analysis by use of English and country of birth (Tables 9.4 and 9.8)

A score describing the number of contexts in which children use a language other than English was developed in the section on language in Chapter 7. Among south Asian children, there is a fairly strong relationship between the test scores and the use of languages other than English according to this measure. The greater the extent to which children use other languages, the lower they score, not only in reading, writing and verbal reasoning, but also in maths and in non-verbal reasoning. These differences persist to the end of the second year. As stated, this finding could be misleading. The relationship probably arises because among this generation some south Asian children (often recent migrants) both speak one or more Asian language and have inadequate English. In the longer term there is no known reason why Asian children should not retain Asian languages but also acquire excellent English.

Only a small number of the tested Bangladeshi children were born in Britain, but it seems that they achieved considerably higher scores than the majority who were born outside Britain. This provides some indication that the low scores of Bangladeshis are associated with the recent arrival of the group. For the other Asian groups, the differences between those born in Britain and elsewhere are small and not of any importance.

Analysis by family background (Tables 9.5 and 9.9)

Very large and consistent differences in the first-year scores are shown according to the socio-economic group of the family. Families where no parent had been in work in the past five years fit at the bottom end of the scale, below the families of unskilled manual workers. The range, in the case of the reading scores, is from -0.54 for children belonging to this 'underclass' to 0.78 for children of professional or managerial parents, and the pattern shown for the maths scores is very similar. However, the two extreme groups are very different from the others, so that the range between the children of unskilled manual workers and those whose parents were in white collar jobs is not very great (from -0.19 to 0.22 in the case of the reading scores). The reasoning scores

vary between socio-economic groups in much the same way as the maths and reading scores.

Analysis of the first-year scores by the highest qualification of the parents shows that the minority of children (15 per cent of the total) whose parents had professional or tertiary academic qualifications achieved substantially higher scores than the rest. Among those whose parents had lower qualifications, there is not much difference according to what those qualifications are, but children of parents with no qualifications score slightly lower than those whose parents have a school leaving qualification, apprenticeship or lower job qualification.

The socio-economic group and qualifications of parents are closely inter-related, but it is clear that socio-economic group is of much greater importance as an influence on children's test scores. First, the relationship with socio-economic group is much stronger than with parents' qualifications. Second, for that 85 per cent of children whose parents do not have professional or tertiary academic qualifications, the level of lower qualifications is not significantly related to test scores. Third, when test scores are analyzed by socio-economic group and parents' qualifications *in combination* (Table 9.11) differences *within* each socio-economic group according to the level of parents' qualifications are shown to be generally small.

Overall, there is not much difference in the first-year test scores between children in single-parent and two-parent households, and the small difference that does appear may well be associated with social class differences. From these findings it looks as though there is no association between the rather low test scores achieved by children of West Indian origin and the fact that a high proportion of them belong to single-parent households. Whether or not the parents are working is quite strongly associated with the children's test scores in the case of both single-parent and two-parent families. In the case of children in two-parent families, those with one parent working full-time and one working part-time obtain the highest scores; in the case of children in single-parent families, those whose parent is working full-time obtain the highest scores.

As might be expected, since these patterns are the result of persisting underlying relationships, they remain much the same at the end of the second year. In particular, there is no indication of any

narrowing of the differences between children from families belonging to different socio-economic groups.

***Effect of ethnic group and socio-economic group in combination
(Table 9.10)***

In their study of a London cohort, Barbara Maughan and her colleagues found that

for white pupils, lower status occupational backgrounds and eligibility for free school meals were both associated with lower reading levels. In the black [Afro-Caribbean] groups, these associations were much less consistent, children in the more socially disadvantaged subgroups having higher reading scores. This difference is likely to stem from the heavy concentration of black adults in low-paid and relatively unskilled jobs, often quite incompatible with their previous work experience or qualifications.¹

This finding does not seem to be repeated in the present study, though we cannot be sure, because there are only 110 children of West Indian origin (the group highlighted in the Maughan study) who were tested and can also be classified by socio-economic group of parents, and none of these have professional or managerial parents. However, the pattern for West Indian children is quite consistent with the theory that their test scores vary according to socio-economic group in much the same way as for white children. Among south Asian children as a whole, there is a clear relationship between test scores and socio-economic group on much the same pattern as for white children. However, among Moslems (largely a sub-group of the south Asians) there seems to be little relationship of this kind, except that the seven Moslem children in the professional or managerial group achieve clearly higher scores than the rest.

From these findings, it seems to be safe to regard socio-economic group as a variable having a comparable significance for the different ethnic groups. The difference between the results of the inner London study and the present one on this point is probably a reflection of change over time (the PSI children are more than ten years younger than those studied by Maughan and her colleagues).

Conclusions

In Chapter 5 we emphasised the very large differences between the study schools in terms of the social and ethnic background of the children entering them; these differences are all the more striking in that the study covers only a part of the full range of state schools, since those containing few members of ethnic minority groups and those outside urban areas were excluded. In the present chapter we have seen that there are correspondingly large differences in the test scores of the children entering the study schools, and that these test scores vary widely according to the children's social and ethnic background. The preliminary indications are that there are, nevertheless, important differences between the second-year test scores in different schools, after discounting the differences between the first-year scores. A full analysis of these differences will be pursued in the next chapter.

The findings confirm other sources in showing that West Indian boys tend to under-achieve, whereas West Indian girls achieve average scores at the end of the second year. It seems that under-achievement among West Indians is not explained by the relatively high proportion of single-parent families within this group. The ethnic group achieving the lowest scores over the first two years is Bangladeshis. There is some evidence that this is connected with recency of arrival. For each of the other south Asian groups the test scores are only slightly below average. Among south Asians as a whole, there is an association between use of languages other than English and low test scores, but it is probably lack of English among some Asian children rather than bilingualism which is the important factor here.

There is little or no indication that the range between social class or ethnic groups has narrowed over the two-year period, or therefore that low-scoring groups have improved their relative position. In particular, there is no indication that the children for whom English is a second language have tended to catch up in terms of reading skills.

Note

1. Maughan et al. (1985), p118.