CHAPTER 8

Summary

It has been possible to use the data collected in the first phase of the research carried out by the *International Project for the Evaluation of Educational Achievement (IEA)* to examine problems of school organisation where there is considerable diversity of practice between systems. It would be difficult to examine some of these problems directly by experiment, for reasons that are plain enough. But where diversity of practice already exists across countries, it is possible to compare practices, each of which is operating in its natural setting, i.e. within the context of the philosophy, traditions and attitudes inherent in its genesis. It is obvious that these variables which are of extreme importance in education would be extremely difficult, not to say impossible, to control in a specially designed experiment.

The IEA has constructed international mathematics tests and administered them to representative samples of students from four populations in full time schooling: (a) all 13-year-olds, (b) all students in the grade where most 13-year-olds are to be found, (c) all pre-university mathematics students and (d) all pre-university non-mathematics students. Questionnaires to collect background information were also constructed and administered to the students tested, their mathematics teachers and their school principals. The data were filed on to magnetic tape and data analysis was carried out in the University of Chicago Computation Center. The data presented in this monograph have been culled from the IEA data.

The first practice to be examined was that of retentivity—the inverse drop-out rate of a system of education (see Chapter 5). The proportion of an age group still in school in the pre-university year varied for those students studying mathematics from four percent in Belgium to eighteen percent in the United States and for those not studying mathematics from three percent in the Netherlands to fifty-two percent in the United States.

The *average* level of mathematics performance of pre-university students is lower in those countries with larger percentage of an age group still in school at the pre-university level. This is true for both students studying mathematics and those not. However, the performance of the best students is much the same in all systems. However, when the achievement "yield" (mean score multiplied by the proportion of an age group in school) of the pre-university students is examined, it can be seen that by increasing the retentivity of a school system, it is possible for a system to have both a high overall yield and an undiminished élite yield. Germany and Belgium have relatively high yields at the 13-year-old grade level and relatively low yields at the pre-university level.

These facts are of interest particularly in those European systems of education where the possibility of increasing retentivity is being examined and where many strong rearguard actions are being fought mainly concerning the maintenance of academic standards. In future research, it should be possible not only to refine the measurement of "acquired yield" and indicate this in various subject areas, but also to compare "acquired yield" with "required yield" (cf. Dahllöf, 1963). The final decision of whether or not to increase the retentivity of a system will be based on economic, political and many other factors.

The second set of practices to be examined concerned differentiation—inter-school grouping, and within the field of intra-school grouping, the practices of ability grouping and age versus grade promotion (see Chapter 6). Unfortunately, no adequate measure of the extent of inter-school grouping exists (in future research, suitable measures should be created; a possible lead might be the coding used for School type Selectivity in Pidgeon *et al.*, 1967). However, a scrutiny of the data available for 13-year-olds and equivalent grade populations suggests a positive relationship between the standard deviations of scores and inter-school grouping. Grade promotion systems have smaller standard deviations than age promotion systems; furthermore, the greater the degree of grade repeating, the smaller the standard deviation. The more ability grouping practised in a system, the larger the standard deviation of scores. However, when the amount of ability grouping practised was partialled out of the relationship between grading and the standard deviation of scores, there was no relationship for the 13-year-olds' scores (i.e., those who, in grade systems, are spread across several grades).

Thus, inter- and intra-school ability grouping is associated with large standard deviations. From other knowledge, it would seem that it is the lower social groups (culturally disadvantaged children) who are mainly responsible for the wide standard deviation by having low scores. In a non-differentiated system, they tend to score higher, thus reducing the size of the standard deviation. Although the range of scores required within a society must be determined on other than purely educational grounds by that society, there are strong arguments for the creation of a non-differentiated system, if the assumption is made that it is the duty of society to give every opportunity to each child to develop to his maximum. It is, however, pointed out that the problem of change in the area of differentiation is not merely that of taking an administrative decision for change, but that of changing the attitudes, particularly of the teachers, within the society—*de jure* abolition of a practice does not mean that *de facto* it will not exist (cf., inter-school grouping in Japan). Furthermore, it should be realised that if inter-school grouping is abolished, but intra-school grouping remains, the standard deviation of achievement scores will not be much reduced.

The third practice to be examined was that of specialization (the number of subjects studied) in the pre-university year (see Chapter 7). The conclusion is that specialization, in the sense of restricting the number of subjects studied in the pre-university year, is not necessarily related to higher scores in mathematics.

The fourth practice was that of mandatory age of entry to school (see Chapter 7). Table 7.3 shows that there is not much to choose between entry at 5 years of age and entry at 6 years of age but that lower scores at 13 years of age are associated with entry at 7 years of age. When the performance of 13-year-old students from different social groups is examined, it would appear that students from higher social groups benefit more from early entry to school than do students from lower social groups, but it is difficult to draw firm conclusions, because of the heterogeneity of scores within each of the age of entry groups.

It is clear that to make the mandatory age of entry to school earlier (e.g. from six to five) will not in itself improve performance; it is what happens in that extra year which is important. This is particularly true for the children of blue-collar workers. It is the qualitative differences which must now be the subject of more systematic research.

An examination of other variables likely to account for differences between countries in the mathematics scores of 13-year-olds revealed the importance of the student's opportunity to learn the mathematics involved in the tests (as rated by the mathematics teachers). This is related to some extent to the qualitative differences mentioned in the paragraph above. It will be of particular interest to mathematics educators to examine the statistics of each item in each of the countries and to consider why 13-year-olds in some countries can perform well on the item while their counterparts in other countries perform only poorly.

Of the other variables examined, important ones seem to be the pre-service training of the teachers and the number of hours of total homework (not just mathematics homework).

Although the first object of any inquiry of this kind must be to find evidence of association there is a further, more difficult, question. When evidence of association has been found how is it to be interpreted? Evidence of association is necessary if causal relations are to be inferred, but it is not enough. When we find an association between the amount of rainfall and the growth of crops we infer that it is the rainfall that causes the growth and not vice-versa. But when we find an association between interest in mathematics and performance in mathematics there may be a difference of opinion whether it is the interest that promotes the performance or the performance that promotes the interest.

In this study the author has presented the evidence of association, and has gone on to use the evidence to make those inferences which seem to him most likely. He recognises that in the last resort the interpretation must depend upon memory, introspection, and testimony and these may differ from one interpreter to another. These are grounds for caution in interpretation. They are not grounds for refraining from the attempt to interpret.

This study, and the parent study (Husén *et al.*, 1967), are first attempts at quantitative international surveys of educational achievement. At the outset many novel problems of measurement, representation and control were encountered. In the later stages there were problems of interpretation. It is to be expected that as time goes on more progress will be made in dealing with these difficulties, and that some of the conclusions reached on the present evidence may need revision as better evidence accumulates. But it may not be unduly sanguine to hope that some, at any rate, of the conclusions will stand.