ACTA UNIVERSITATIS STOCKHOLMIENSIS
Stockholm Studies in Educational Psychology 15

SCHOOL ORGANIZATION AND STUDENT ACHIEVEMENT

A Study Based on Achievement in Mathematics in Twelve Countries

BY

NEVILLE POSTLETHWAITE

ALMQVIST & WIKSELL
STOCKHOLM

CONTENTS

TABLES

Chapter 5

Chapter 6

Chapter 7

FIGURES

Chapter 4

Chapter 5

Chapter 6

APPENDIX

CHAPTER I

Frame of Reference

Comparative education as a discipline is concerned with the study of cross-national or cross-cultural variability in the domain of education. Until the beginning of the 1950's it consisted mainly of separate descriptions of various systems of education; many of the comparative education text-books of that time consisted, with few exceptions, of a collection of chapters, each describing a particular nation's system of education. In the fifties, parts of one system were placed side by side with similar parts of another system and were described in more detail than when the systems as a whole were compared. Bereday (1964) has called this the "juxtaposition" stage. The emphasis has been on the exchange and collation of descriptive material. International agencies such as *UNESCO, O.E.C.D., I.B.E.* and the Council of Europe's *Council for Cultural Co-operation,* have helped to intensify this exchange and collation, with the result that there exists a wealth of data relating to different patterns of educational organisation, curricula and teaching methods. However, where any analyses of these data have been undertaken, these have been of a qualitative nature and usually within countries.

It has become increasingly evident that formal education plays an important part in the social, economic and technological development of a country; at the same time, the scarcity of resources has made it impossible both in developed and developing countries to satisfy the growing demand for educational expansion and this, in turn, has underlined the need for a critical inquiry and re-appraisal of some of the educational practices in existence today. Anderson (1961) has indicated the need to introduce into comparative educational studies established procedures of research and quantitative assessment, so as to gather information not only about the "efficiency" of various types of educational systems, but also about the "efficiency" of various educational practices within them. Bereday (1964) too, has emphasised the need for an analytic (qualitative and quantitative) stage in comparative education—the post-juxtaposition stage. As a result of such cross-cultural analyses it should be possible to draw conclusions on the basis of inductive reasoning.

The efficiency (in terms of optimum production of learning—both cognitive and non-cognitive) of schools in various nations is attacked and defended usually without solid evidence to support the claims of either attackers or defenders,[1] with the result that policy is often made on the basis of assumption and impressionistic and incomplete evidence. A United States admiral, in a widely publicised article in 1965, contended that one school year in the United States would be worth only two-thirds of a school year in Europe, but, as yet, no evidence has been gathered by which this impressionistic statement can be confirmed or rejected. The type of statistics which have so far been collated and classified concern "input" variables to the school system (e.g. statistics concerning teachers, buildings, financial expenditure per student, curricula, etc.) but no systematic measures of qualitative "outcomes" have been made (cf. Harbison and Myers, 1964).

Thus, in order to examine the "efficiency" according to certain criteria of systems as a whole, or of particular educational practices, it is necessary to have measures of the "outcomes" of the various systems. This implies that internationally valid cognitive and non-cognitive measures (in the form of tests, attitude scales and questionnaires) are used, so that comparable data are obtained about a number of educational systems at the same time.

Such data are of special value:

1. when one wishes to study the relationship between certain variations in educational practice and educational achievement, but the practices and school structures one wishes to compare are not well represented within a single country,
2. when it is desirable to test the generality or universality of a relationship that has been found in some country.

One illustration of the former would be an inquiry into the relationship between the age of commencing formal schooling and subsequent achievement. How does achievement at, say, age thirteen, compare for students who entered into formal schooling at age five, age six, or age seven? The uniformity of practice within a single

[1] Cf. Carnegie Quarterly, Volume XIV, No. 2, 1966: "The Gross Educational Product: How Much are Students Learning?"

country almost precludes any study of this question within national boundaries. It would be extremely difficult, if not impracticable, to set up an experimental situation within a single country. Furthermore, it would necessitate changing many of the cultural assumptions and values held by teachers, students, parents and society for the various experimental situations. Such variety, however, already exists internationally and an international study would reveal the diversity of practice in different countries and make data on this point readily available.

An illustration of the second type of relationship is the allegation that boys do better than girls in certain subjects. Is this a general phenomenon, or is it limited to certain countries? If the latter is the case, what are the characteristics of the cultural patterns and of the educational systems in which boys do better and of those in which girls do better?

Thus, an international study of education must centre on the kinds of questions that can be answered best (or solely) by comparisons of the achievement of students in different countries, and that can be answered poorly, if at all, by studies of students within a single country. The school systems of the world represent a series of environments in which human beings learn, and as a group are much more varied and contain far greater differences than can usually be found or created in any one system. Thus, educational quasi-laboratory situations exist in which many of the important questions concerning human learning can be studied objectively, though there is still a great deal of difficult work involved in specifying such environments with reasonable accuracy and in comparable and meaningful ways.

The design of the international research study reported here is of the survey type using random probability sampling techniques. As implied earlier on, the survey approach has the advantage of examining practices as they exist, and with the surrounding philosophies and values concerning those practices held by the students, teachers, parents, and other members of society. Degrees of association between certain independent (input) and dependent (output) variables can be measured, as well as between certain of the independent or dependent variables themselves. Although it is more difficult to infer cause and effect relationships than in a controlled experiment, it can be argued that it is extremely difficult to set up a controlled experiment in the educational field for examining certain problems (cf. Carroll, 1963, Kish, 1965). For example, for a controlled experiment involving an examination of streaming, it is important that teachers teaching streamed classes should believe in the principle of streaming and vice-versa for those teaching non-streamed classes. In practice, this is difficult. On the other hand, it is well known that there will always be teachers of differing philosophies teaching both sort of classes. A survey research can look at the situation as it exists and evaluate streaming versus non-streaming in their various contexts. This is obviously of more value than an examination of the problem in a artificially set-up experiment. However, it must be borne in mind that only experimental studies allow conclusions of cause and effect relationships. Any notion of cause and effect from survey research is strictly inferential.

The present study has drawn from the data gathered by the IEA Project (see Husén *et al.*, 1967) where educational research centres or institutes from twelve countries: Australia, Belgium, England, Federal Republic of Germany, Finland, France, Israel, Japan, the Netherlands, Scotland, Sweden and the United States, participated in a cross-national study of a comparison of the outcomes of mathematics instruction. A short account of the history of the project, the International Project for the Evaluation of Educational Achievement (IEA), as well as the problems of choosing comparable populations for testing and sampling them and how these were overcome is given in Chapter 2. The instrument construction, data collection and data processing are described in Chapter 3. Chapter 4 describes certain of the independent variables used in the study presented here.

The educational practices chosen for study in the present book are those where wide variation occurs between educational systems. They, therefore, concern data where a study is made of the relationship between certain variations in educational practice and educational achievement. Previous research and the results of this study are given in detail in Chapters 5, 6 and 7. A brief introduction is given here of the problems associated with each of the practices examined. However, it must be made quite clear that these problems are being examined in terms of only *one* aspect of student achievement—mathematics. Whether the results would be the same in other subject areas is a matter for future research.

Retentivity[2]

Retentivity is a term used to describe the proportion of an age group being retained in full-time schooling in a system to the end of secondary schooling. Thus, the United States system, since it retains nearly three-quarters of an age group in school through to the end of twelfth grade, is described as a "highly retentive" system, whereas England, retaining only twelve percent of an age group, could, in 1964, be described as a system with "low retentivity".

In the United States and Japan, which are highly retentive systems, there would appear to be a deliberate policy of encouraging as many students as possible to continue through to the end of secondary schooling. In many European countries, there has been a policy of gradually selecting out a small élite which has been allowed to continue through to the pre-university year. Theoretically, of course, each child is allowed to continue through, but usually on condition that various academic (and selective) hurdles are overcome. In the last decade in Europe, as steps have been taken to broaden the opportunities for secondary and higher education, the objection has frequently been raised that, if more students are allowed through either to the pre-university year, or to the university, this will mean a "lowering of standards". Unfortunately, when asked for an operational definition of "standards", those who use the term are either at a loss to supply one, or suggest, that "standards" refer to the minimum requirement for a "pass" mark that has emerged over the years (cf. Husén, 1966).

By the use of internationally valid mathematics tests, it is possible to compare the outcomes of students both studying and not studying mathematics in the pre-university year. It is possible to compare the outcomes from various points of view. First, it is possible to compare the average performance; it is often asserted that the "standard" of performance of the students in the pre-university year in the European low retentivity systems is higher than that of the United States twelfth-graders—is this true or not? Secondly, it is possible to examine the relative performances of students at different parts of the distribution of scores in each system. Thus, for example, how do the top five percent in school in England compare with the top five

[2] Some aspects of this problem have also been taken up in the IEA international publication (Husén *et al.*, 1967) by the present author and others.

percent in school in the United States? Is it true that if more students are allowed through to the pre-university year, this will mean a lowering of "standards" for the "best" students? Since the degree of retentivity varies greatly from country to country, it is obvious that a comparison of international percentiles referring to the composite distribution of pre-university mathematics (and separately non-mathematics) students is not fair to the highly retentive systems. Therefore, it is necessary to go one step further and calculate the proportions of a *total age group* reaching various levels of achievement. It can be appreciated that a higher proportion of students in full-time schooling in a low retentive system are likely to reach, say, the international 95th percentile than in a high retentive system, but that when the same two countries are compared in terms of the *total age group* reaching the 95th percentile, the reverse may be true. Calculating the proportion of a total age group reaching certain "standards" (in terms of international percentiles) introduces the concept of "how many students are brought how far" in a particular system. It is possible to develop this line of thought and calculate an "achievement yield" of particular groups of students. This takes into account the percentage of an age group reaching a particular level of achievement, and is not simply a comparison of means between countries irrespective of the differing percentages of an age group making up the population being compared. Furthermore, it is possible to compare the "increase" in yield between a point where one hundred percent of an age group are in school (in this study, 13-year-olds) and the pre-university year. Ideally, it would be desirable to measure the "total yield" of achievement of a system. This, however, would require measuring achievement of all those dropping-out of school at the points at which they drop out. Another approach would be the longitudinal, measuring student accomplishment at the beginning and at the end of a given school year or stage.

Differentiation

Differentiation is a term used to describe the policy of grouping students by some particular criterion into different schools or into different classes within schools (Husén, 1962 *a*). In selective systems of education, students are separated, usually somewhere between the ages of ten and twelve, on the basis of ability and/or achievement,

into separate school types. The more able students go to a selective academic school (*grammar school, lycée, Gymnasium,* etc.) and the others continue in a form of elementary school (*modern school, école primaire, Volksschule,* etc.). This type of differentiation is sometimes known as "organizational differentiation" or "inter-school grouping". When a similar form of grouping is practised within schools (grouping students by ability or achievement into classes) this is sometimes known as "educational differentiation" or "intraschool grouping".

In the twelve countries participating in the IEA study, there was more diversity between countries than within any one country in the forms and amounts of differentiation employed. Previous studies (see Chapter 6) have implied that the more differentiation practised either within a system or within a school the larger will be the range of achievement; at the same time, there is other evidence (Marklund, 1962, Svensson, 1962, and Husén *et al.,* 1967) to suggest that the mean scores of "bright" students are, in the long run, much the same whether they have been subjected to the policy of differentiation or not, but that "duller" students achieve more when in a non-differentiated system of education or school than in a differentiated one. However, in any system of education, it can thus be argued that it is the achievement of one hundred percent of an age group which is as important, if not more important, than the achievement of a small élite. It is, therefore, of interest to examine the range of scores on an achievement test in relation to the amount of inter- and intra-school grouping practised in various systems of education. If it is true that larger ranges of scores are associated with the amount of differentiation practised, then educational policy makers, planners and administrators should be aware of this when planning policy. It is also of interest to know the relationship of inter- and intra-school grouping, both together and separately, with variability of achievement. For example, if it is planned to change from a selective to a comprehensive system of education, but it is expected that intra-school grouping will be practised in the comprehensive school, then what will be the approximate change in the variability of achievement? Alternatively, if intra-school grouping is not practised, then what might be the change in the range of achievement of a year group?

Related to the aspects of inter- and intra-school grouping is that of grade promotion versus age promotion. Some systems of education insist on students reaching a certain level of achievement before being allowed to progress to the next grade; this results in certain proportions of an age group being one or two grades behind the majority of their contemporaries. Other systems allow a total age group to progress as an age group through the school. It is to be expected that a grade system will have a smaller range of achievement within any one grade, but a larger range over any one age group. On the other hand, there will be an interaction effect between the age-grade progression (the promotion system which is in itself a form of grouping) and the amount of intra-school grouping practised within a grade or age group. Is it possible, for example, that within one year group in England with age grouping, but with streaming within an age group, the range of scores will be larger than in a system with grade grouping but no streaming?

The diversity of differentiation practised in the IEA study has made it possible for these questions to be examined to some extent, i.e., the relationships between various forms and degrees of differentiation and the standard deviation of scores. The results are to be found in Chapter 6.

Specialization and Age of Entry to School

Two other aspects of school organization where diversity exists between systems but not within systems are those of *specialization* (the practice of gradually dropping subjects or not dropping them, so that by the pre-university year only a few subjects are studied, or as many as in the early years of secondary school) and mandatory age of entry to school.

In England and Scotland, students in the penultimate and ultimate secondary school years study an average of three of four subjects only; in the United States, students in twelfth grade take three or four "solids", but it is theoretically possible that in eleventh grade they could have taken three different "solids". In many Western European countries nine or more subjects are studied through to the end of secondary school. In those countries where only few subjects are studied, there has been much discussion as to whether students should study more subjects. Those in favour of studying more subjects have pointed out that it is early enough to begin specializing at the university, and that at school a more general education should

be given, since with the speed of technological change in today's world, many persons will have to be retrained several times in their lifetimes for new jobs, many of which do not even exist today. Furthermore, the fact that specialization takes place in the last years of school has a backwash effect, with the result that many students who drop-out of school before reaching the pre-university year have already dropped some subjects and in some cases are studying clusters of subjects which are arts or science biassed. Those in favour of specialization argue that it is important to concentrate on only a few subjects, since this keeps up "standards" of achievement in the pre-university year, that the universities require this specialization and that by studying a subject in depth, students are more capable of appreciating higher thought processes and that their achievement will be of higher level than those who study more subjects.

Thus it is of interest to compare the achievement of pre-university students from different countries according to whether specialization is practised or not. In general, within a country where, on average, few subjects are studied, it is difficult to examine the problem, since it is the "brighter" students who tend to study more than the average number of subjects. It should not be forgotten, however, that there are difficulties in making a straight comparison between countries on this variable, since differences between the groups of students exist which are of importance, notably that the average age of terminating the pre-university year is different from country to country and that the percentages of an age group going through to the pre-university year also differ.

England and Scotland have a mandatory *age of entry* to school of five years, Sweden and Finland of seven years and the other countries in the IEA study of six years. The median age of entry differs slightly from the mandatory age, but not sufficiently to require a different categorization in terms of the average length of schooling up to a particular point later in the systems. This particular diversity in educational practice has been mentioned earlier in this chapter as an illustration of the advantages of international educational research over national research. However, within some countries there is some small variation and interesting national studies have been carried out (Pidgeon, 1965). Those who support an earlier age of entry to school maintain that early entry makes early learning possible and that students who enter earlier will learn more than those who enter later; furthermore, it is easier for them to learn social adjustment to their peers at an earlier age, and that for "culturally deprived" children the deprivation can best be compensated by bringing the children to school earlier.

In this study, it is possible to compare the achievement of 13-year-olds in twelve countries and relate this to the mandatory age of entry to school. It is also possible to compare the relative achievement in various socio-economic status groups on the same variable. Do, for example, low socio-economic status group 13-year-old students have higher achievement scores in those countries where they begin school at five years of age than in those countries where they begin at six or seven?

It has been shown that when pre-university students' mathematics scores are adjusted for differences in age and retentivity in the different systems, the differences in scores between countries are much the same as at the 13-year-old stage (Husén *et al.*, 1967). This being so, it is interesting to add other features of school organization to that of age of entry and examine to what extent school organizational differences can account for the differences in score. It is not likely that this will be very great, since, on the basis of previous knowledge (Peaker, 1967) it is known that school and teacher variables account for a relatively small amount of the variance of scores. Nevertheless, it is of interest to those concerned with school organization to be aware of the effects of their policies.

Summary

Comparative education as a discipline has now advanced to the stage where it is necessary to carry out cross-national empirical studies of not only the input (independent) variables to systems of education, but also the "outcomes" (dependent variables) of the systems. Data collected in international studies are of special value:

1. when one wishes to study the relationship between certain variations in educational practice and educational achievement, but the practices one wishes to compare are not well represented within a single country.
2. when it is desirable to test the generality or universality of a relationship that has been found in some country.

Furthermore, international surveys of educational systems have certain advantages over small-scale controlled experimental studies. First, they involve replication and secondly, the practices being studied exist in their natural contexts with all the concomitant philosophies and value systems as they exist in practice. In a controlled experiment, it is often extremely difficult to control variables such as teacher attitude (philosophy) and once it is carried out, it requires replication.

The International Project for the Evaluation of Educational Achievement (IEA) has recently undertaken a study of mathematics achievement in twelve different school systems (Husén *et al.*, 1967). The data presented in this book come from the IEA study. The educational practices examined are those where there is considerable diversity between countries and considerable uniformity within countries.

The first practice is that concerning the differing proportions of an age group continuing through to the pre-university year (retentivity). It is intended to examine the differences in "standards" of performance associated with differing degrees of retentivity in terms of average performance, fixed international standards performance and "yield", the latter being a measure of how many students in certain defined populations are brought how far in terms of achievement within any one system. These results are reported in Chapter 5.

The second practice is that of differentiation. Students are differentiated into different school types (inter-school grouping) and into different groups within schools (intra-school grouping) to differing degrees formally on the basis of ability and/or achievement. It is possible to examine the association between these two forms of differentiation and the spread of achievement scores. Further, practices differ between countries as to how students are grouped in connection with promotion policies; some countries have a system of grade promotion and others a system of age promotion. It is possible to examine the spread of scores in connection with these forms of grouping and in turn the relation between these two and the relation between spread of achievement scores and intra-school grouping. These results are reported in Chapter 6.

The third and fourth practices concern the number of subjects studied in the pre-university year and the mandatory age of entry to school. It is possible to compare the mathematics scores of students from countries where nine or more subjects are studied with students' scores from countries where only three or four subjects are studied. The mandatory age of entry to school ranges from five to seven years of age in the countries participating in the IEA project. Is earlier mandatory age of entry to school associated with higher achievement scores at age 13 in general, or only for some social groups? Are there other school organizational features which account for differences in score between countries at the 13-year-old level? These results are reported in Chapter 7.

All of these problems are those on which some light can be shed from the results on an international study, but which would be difficult to examine within a single nation. However, it must be remembered that these results refer only to mathematics achievement; it would require further research to check these results in other subject areas.

CHAPTER 2

IEA, Populations and Sampling

International Project for the Evaluation of Educational Achievement (IEA)

The data used in this study were collected by the *International Project for the Evaulation of Educational Achievement (IEA)*, and since IEA is the first large-scale international educational research project of its kind, it would seem appropriate to describe briefly its history, structure and mode of operation. A detailed report of the IEA project is given in Husén *et al.* (1967).

In the middle fifties, groups of educators and educational researchers from different countries had met at places like the UNESCO Institute for Education, Hamburg, to examine problems such as those concerned with school structures and organization, selection processes, examinations and failure in school. Two important publications emerging from some of these meetings were edited by Hotyat (1962) and Wall (1962). Throughout these meetings there was a growing awareness of the need to establish evaluation techniques which would be valid cross-nationally. At the same time, more or less independently of each other, several researchers in the United States (Anderson, Bloom and Foshay) began to consider the possibilities of undertaking such research.

In 1958, researchers from several countries came together at a meeting in Eltham, England, chaired by Dr. W. D. Wall of the National Foundation for Educational Research in England and Wales, and also at the UNESCO Institute for Education in Hamburg. At those meetings it was decided to carry out a pilot study to discover if an international research project would be administratively possible and if the results could be expected to be meaningful. Research Centres from Belgium, England, Finland, France, Germany, Israel, Poland, Scotland, Sweden, Switzerland, the United States and Yugoslavia took part. A strategic target population in those countries was the children of age 13:0 to 13:11, since this was the last point where practically all of an age group were still in school in all countries. In most cases, children of schools or areas which were known to be close to the national mean and standard deviation were tested, and thus, there was no strict probability sample. In all, 9,918 students spread over eight languages were administered tests (a total of 120 items) of reading comprehension, mathematics, science, geography, and non-verbal ability. The venture proved to be successful. Foshay *et al.* (1962) have presented some of the results of this study in a monograph.

At a meeting at the Unesco Institute for Education, Hamburg, in June 1960 it was decided to embark on a cross-national study in one subject area, where several populations within secondary education would be sampled using random probability sampling techniques and where specific testing instruments would be specially constructed. This first carefully designed study in one subject area would be known as Phase I and it was hoped subsequently to embark on further phases.

The subject chosen for the first phase of the project was mathematics. The primary reason for this choice was that most countries involved in the project were concerned with improving their scientific and technical education, at the basis of which lies the learning of mathematics. Secondly, many recent national and international surveys (as carried out by the National Science Foundation in the United States and O.E.C.D. in Europe) have re-examined the curricula and the methods of teaching mathematics and various higher branches of mathematics. Thirdly, the so-called "new mathematics" has been introduced to varying degrees in some of the participating countries. Fourthly, since the symbols of arithmetic and mathematics are, with trifling exceptions, international problems of semantics and language would be reduced.

The Research Centres which committed themselves to Phase I at the 1960 meeting were from Belgium, England, Finland, France, Israel, Japan, the Netherlands, Scotland, Sweden and the United States. It was in late 1962 and early 1963 that Research Centres from Australia and Germany entered the project. (The main persons involved from each of the Centres as well as consultants are listed in Table A1 in the Appendix). A research grant from the United States Office of Education was received in the summer of 1962 and this

covered the international costs and the United States national costs only. The representatives of the Research Centres from these twelve countries formed themselves into a Council whose main task was to agree on the overall policies of the research work. On average, they met for a week once a year. They elected a Standing Committee of five of their members and their task was, if necessary, to take major decisions between Council meetings on behalf of the Council. Furthermore a Chairman/Technical Director was elected whose task was to attend to the day to day running of the project. He was assisted by a Project Co-ordinator, who was appointed in 1962 and placed in the UNESCO Institute for Education, Hamburg.

In such a project, the lines of communication were long, and it was very important to set deadlines for various stages of the work and to adhere to them. Several languages were represented, and it was decided that the project should be conducted in English, with occasional French translation. Although there were some misunderstandings, they were fortunately rare. Lessons were learned from experience and improvements in the mode of operation were continually undertaken. A list of "lessons learned" is given in Chapter 2 of Volume I of the international publication.

Consultants were employed in the areas of mathematics education test construction and sampling, and these consultants attended all Council meetings as well as special group-work meetings, which were sometimes held between Council meetings. A great deal of group work was also carried out at Council meetings; thus, for example, further work on mathematics test construction, attitude scale construction, questionnaire construction, formulation of hypotheses and sampling took place in the early meetings. After the full testing, all members helped in writing up the outcomes of the testing of hypotheses.

In its turn, the National Centre, although using most of its own staff on the national work involved in the project, sometimes used sampling consultants. At the content analysis stage at the beginning of the project, the National Centre had to organize national committees of mathematics educators and at the coding and punching stage, they often had to employ extra coders (mostly university students).

The data were put onto magnetic tape at the University of Chicago Computation Center. Needless to say, with approximately fifty million pieces of information, this study could never have been completed without the use of a computer. That the whole project (mathematics phase) was completed within four years, even with the help of a computer, was, in itself, an enormous achievement—the work on content analysis was begun at the beginning of 1962 and the final research reports were completed at the end of 1965; this success was due to the dedication, enthusiasm and ability of all the educational researchers concerned. The data on the master and working tapes at the University of Chicago Computation Center will form a data bank which can be used by qualified research workers. A Data Bank Manual has been prepared by Richard M. Wolf (1967).[1]

The IEA Council has decided to embark on a second major phase where testing in other subject areas will be undertaken, and the frame of reference of the research will be extended in terms of the various psychological, social, cultural and economic forces involved in the process of education.

Populations Tested

One of the most difficult problems in a comparative study of this nature is deciding which populations in the different countries are, in fact, for one's purpose comparable. The pilot project (Foshay *et al.*, 1962) had focussed on the educational attainments of 13-year-olds. This group has the merit of being the highest age level at which, by law, all children are supposed to be attending school in most countries with a tradition of universal education. The 13-year-old group had distinct advantages, therefore, for an assessment of the educational standard reached by an approximation of a total age group of each country and was thus selected. Although this group was chronologically comparable, there were difficulties in that there is a wide variation between countries as to the grades in which 13-year-olds are to be found. In some countries, its members were nearly all in the same grade, while in other countries, because of retardation or acceleration policies, they could be spread over several grades. For example, in England, Scotland and Japan, approximately ninety-

[1] A copy of the Data Bank Manual can be obtained upon request to: IEA Coordinator c/o Unesco Institute for Education, 2 Hamburg 13, Feldbrunnenstr. 70, Federal Republic of Germany.

nine percent of a year group are to be found within the same grade, whereas in Belgium, for example, twenty-nine percent of 13-year-olds are retarded by one, two or three years. In the latter case, it was thought to be difficult in the testing programme to have all of these children brought from the different classes, and in certain cases, different schools, to the testing session. It was therefore decided to allow Research Centres to award a notional zero score to those children whom they considered to be so retarded as to be unable to attempt any of the questions in the tests. However, in most cases, all students of this age range were, in fact, tested.

A second population, which is the complement of the first population, is that consisting of all students at the *educational* level (grade level) typical of the 13-year-olds in each country. This, then, is an educational level population designed to correspond in general, to the age represented in the first population. The 13-year-old age population was designated Population 1a, and the 13-year-old grade group was designated Population 1b.

The grade group, containing the majority of 13-year-olds will, of course, be different according to the time of year chosen for testing. Take a hypothetical example of two year groups: *a*) 13–14 and *b*) 12–13 at the beginning of the school year. Then, further assume that the school year runs from April to March in the next calendar year. Thus, if testing takes place between April and September, the 13-year-old grade group which will be tested will be group *a*, but after September, will be group *b*. To avoid disparity, it was agreed that the tested group would be the grade where the majority of 13-year-olds were to be found within three months of the end of the current school year. It must be pointed out that in almost no country did Populations 1a and 1b represent students at any terminal point. Therefore, their achievements are *not* to be considered indicative of what has been achieved in a rounded-off course of study. They *do*, however, provide a more or less hundred percent attendance base-line against which further learning within the system of secondary education can be measured.

Another group of students who seemed of special interest were those who were just completing the pre-college or pre-university level of education. This represents a major transition point in each educational system and also is the termination of formal schooling in each country. It is also a point which can be said to be that where the "fruits" of education may be assessed. Obviously, however, there are important differences between countries in the composition of these groups. For example, the average age of completing pre-university education ranges from 17 years 2 months in Australia to 19 years 10 months in the Federal Republic of Germany (cf. Chapter 14 in Vol. I, Husén *et al.* 1967). Again the age at which students begin school varies from country to country, and thus the total length of schooling varies. Secondly, it can be argued that the second and third year sixth-former in an English state school is not the equivalent of an American 12th grader or even of a Swedish *studentexamen* student. Apart from different lengths of schooling, the selection process which has taken place in each of these systems is very different in terms of grade-repeaters and drop-outs, and the number or the percent of a year group in this pre-university year also differs from country to country. Thirdly, the number of subjects studied in the pre-university year ranges from an average of three in England to nine or more in some European countries. Thus, there are differences in the structure of this transition point from one country to another, and this must be borne in mind in the interpretation of the results. However, it was decided that the advantages of working at the pre-university major terminal point appeared to outweigh the disadvantages of lack of comparability, so this population was chosen. It was divided into two sub-populations on the basis of the curriculum being followed. One sub-population consisted of those taking mathematics as a major subject. The second group was made up of those who were not taking mathematics or for whom mathematics was a minor and subsidiary part of their programmes. In most cases the two groups belonged to different sections or tracks of the pre-university school.

Between the 13-year-old level and the pre-university year, there are various major terminal points in the school systems—e.g. end of compulsory school ranging, for example, from 14 years in Germany to 16 years in France, Sweden and the United States, and major examination points such as the G.C.E. "O" level in England. Thus, in some countries these populations represented students terminating their education at the intermediate level, and in other countries they represented a kind of half-way point between the lower and the pre-university populations. It was decided that countries could choose the population(s) they wished to test at these in-

termediate points. The following are the formulated definitions of the target populations. As indicated above, it was stated that testing should take place within three months of the end of the academic year. The mathematics tests (see Chapter 3) given to the students in each population are given in parentheses:

Population 1a:

> *All students who are aged between 13:0–13:11 years at the date of testing.* This means that all types of schools with students of this age should participate and be represented according to their proportions of students from the population defined.
> (These students were to be given Mathematics Tests A, B and C.—See page 42.)

Population 1b:

> *All students at the grade level where the majority of students of age 13.0–13.11 are found.*
> (These students were to be given Mathematics Tests A, B and C.)

Intermediate Populations (Optional):

> These target populations were defined by the countries testing at these levels. It was desirable, however, that, where possible, these populations should be taken at points which, if terminal, did not lead to universities or similar institutions of higher learning.
> (These students were to be given Mathematics Tests 3, 4 and 5.)

Population 3:

> *All students who are in the grades (forms) of full time study in schools from which the universities of similar institutions of higher learning normally recruit their students.* These students, in most countries, were in the grades (forms) from which a qualifying examination for the university of similar institution was taken, e.g. *Abitur, Studentexamen, 2^e partie du baccalauréat, Eindexamen, G.C.E "A" level.*

Qualification—This did not include the small proportion going to universities or similar institutions of higher learning via institutions which came under the heading of "Zweiter Bildungsweg", but the proportion of the population had to be known. Population 3 is divided into two parts:

3a: Those studying mathematics as an integral part of their course for their future training, or as part of their pre-university studies, e.g. mathematicians, physicists, engineers, biologists, etc. or all those being examined at that level. (These students were to be given Mathematics Tests 5, 7, 8 and 9.)

3b (highly desirable, but optional): Those studying mathematics as a part (complementary) of their studies and the remainder. (These students were to be given Mathematics Tests 3, 5 and 6).

Where Centres wished to sub-divide any of the above populations for national purposes, they were, of course, allowed to do so.

For purposes of coding, it was then necessary to create "operational groups". For example, in the following section, it can be seen that Groups 1 and 2 form Population 1a, and Groups 1 and 3 form Population 1b. Populations were thus broken down into operational groups as follows:

Definitions of Groups

Group 1 consists of those students aged between 13.0 and 13.11 on the day of testing in the grade (or year group) which contains the majority of students of this age.

Group 2 consists of those students aged between 13.0 and 13.11 on the day of testing who are in grades (year groups) other than that in which the majority of this age are found.

Group 3 consists of the remainder of students in the grade (year group) from which Group 1 is taken.

Group 4—Level 2(i) as operationally defined by National Centres.

Group 5—Level 2(ii) as operationally defined by National Centres.

Group 6—Level 2(iii) as operationally defined by National Centres.

Group 7—Level 3a as operationally defined by National Centres.

Group 8—Level 3b as operationally defined by National Centres.

Group 9 consists of those students who are tested with Level 3a tests, but who are possibly following a course of mathematics which does not *clearly* place them in Level 3a.

Group 0 consists of those students who are tested with Level 3b tests, but who are possibly following a course of mathematics which does not *clearly* place them in Level 3b.

Since the intermediate populations chosen for testing in the various countries vary so much, it was not thought worthwhile making international comparisons, and therefore these populations were left for national analyses and not included in the international analyses (see e.g. Pidgeon, 1967).

Sampling

Sampling Units and Stratification

The main problem in sampling was to secure a representative sample of the particular target populations in each country. Each national research centre appointed a sampling expert for its country. The IEA, on the other hand, decided that it was necessary to have one person who could devote himself more or less continuously to the task of examining the sampling plans for each target population within each country and who would enter into correspondence with the national sampling expert.

Each target population was divided into a sampled population and an excluded population. It was agreed that where there was a small category of schools that, on the one hand, would be very expensive to sample and, on the other, was so small that the results from it would make little difference to the general picture, it could be reasonably excluded. In all cases, the excluded population was negligible, except in Israel, where students who had recently immigrated from under-privileged areas were excluded.

The procedure used for sampling the "sampled population" was that of stratified random probability sampling. The unique merit of probability sampling is that the standard error of the sample as a whole or of any part of it can be determined from the internal evidence of the sample itself. All of the countries used probability sampling, except for the Federal Republic of Germany (represented by only two of the Länder—Hessen and Schleswig-Holstein) which maintained that if a random process of selection of schools was used, many of them would be unco-operative and that it would be better not to use probability sampling, but to make instead a judgement sample from schools known to be co-operative. This was, of course, for the Germans to decide, but it is clear that the internal evidence, in this case, supplies no guarantee of representativeness.

In the United States, the sampling was in three stages, the first stage being a sampling of communities, the second a sampling of schools within the selected communities, and the third a sampling of students within the selected schools. Elsewhere the sampling was in two stages, with schools as the first and students as the second stage. Multi-stage sampling is needed, because it is impracticable to sample students directly in a single stage. But a multi-stage sample is bound to be larger, in terms of students, than a simple (i.e. a single stage) sample giving standard errors of the same size.

Thus, with two stage sampling, and small sampling fractions, the variance of an estimate is

$$\frac{S}{n}+\frac{P}{nk}$$

where n is the number of schools in the sample, k the average number of students selected within each school, S the variance of school means and P the variance of students within schools. The intra-class correlation—i.e. the measure of the extent to which students in the same school resemble each other more than they resemble students in general—is ϱ where $\varrho = S/S+P$.

Consequently,

$$\frac{S}{n}+\frac{P}{nk}=(\overline{k-1}\,\varrho+1)\frac{P+S}{nk}$$

and $(k-1)\,\varrho+1$ is what Kish (1965) calls the Design Effect (Deff). In other words, it is the ratio of the size of the complex sample, in terms of students, to that of the simple equivalent sample. If the standard errors for the complex sample were calculated by applying simple random sampling (s.r.s.) formulae directly they would be too small. The proper values can be obtained by multiplying the s.r.s. estimates by the square root of Deff.

The Design Effect can be reduced by stratifying the schools, which reduces the intra-class correlation. It could be reduced further by stratifying students within schools. Stratifying schools reduces S, and stratifying students reduces P. In this study schools were stratified but the stratification of students was not attempted. In all countries schools were stratified by sex and type, and in some also by (a) geographical or administrative areas, (b) ethnic and religious groups, and (c) rural-urban locality.

Three principles of random selection of students within the schools were proposed:

1. Working through the registers with a constant sampling interval and a random start.
2. Taking in the students whose surnames begin with certain letters of the alphabet.
3. Taking in the students whose birthdays fall on certain days, spread uniformly around the year.

Research Centres were warned that, when the first principle was used, there is sometimes a strong tendency for schools who draw "unlucky" random numbers to ignore them and to choose, by judgement, a "fairer" sample. Often the headteacher replaces what he considers to be "poor" students by "good" students. This method, in fact, was not used. A warning was also given about the second method—i.e. that there may be an association between the initial letter of surnames and ethnic or other groupings within the society. If this was to produce a bias, it should be avoided. Most Centres used the third principle. This is notionally equivalent to re-defining the population so that it consists only of children with particular birthdays. There is no reason to suppose that the reduced population, defined by birthdays, uniformly spread around the year, differs from the complete population. The size of the samples varied according to the population and the country, but the number of students tested for each population varied from approximately 700 to 6000. All in all, the total number of students tested (including intermediate populations) was about 135,000.

Since the school had been used as the sampling unit, it was decided to deal each population sample into four independent sub-samples. The data were coded in terms of sub-samples and put onto the magnetic tape in this way. The splitting of the population into four independent sub-samples had various advantages. The first was that independent estimates could be obtained from each of the four sub-samples and estimates of error from the comparison of these. The second advantage was an administrative one, namely, that the answer sheets for each sub-sample could be shipped separately to Chicago. Thus, if one were lost, three still remained, whereas if all had been shipped together, all might have been lost.

It turned out that Israel and Australia did not test Population 3b and that France and the Netherlands had to be dropped because of several cases of undersampling of schools. The Federal Republic of Germany and Israel did not test Population 1a.

Weighting[2]

The actual sampling fractions differed somewhat from those suggested in the original sampling design handed in by the national sampling experts. The two main reasons accounting for this disparity were (1) the numbers of schools taken into the sample in each stratum were based on national statistics dating back as far as 1960 or 1961, and in 1964 when the testing took place, there were changes in the figures, and (2) in certain cases it was not possible to test all students drawn within schools which had been sampled. In some cases the school refused to cooperate in the study, and it was too late to take an alternate school in terms of the test programme administration within that country. The differences were not great, however, but it was the actual and not the designed sampling factions which were used to obtain the raising (weighting) factors. The weighting of each stratum sub-sample was carried out in such a way that the weighted number of students in each stratum was in exact proportion to the *total* number of students in each stratum. The estimates of error used in reporting the results in this study are those obtained from the comparison of the estimates of each of the four sub-samples. The formula used for weighting was:

$$ {}_1n_i = {}_2n_i = {}_3n_i = {}_4n_i = \frac{nN_i}{4N} = \frac{n_i}{4} $$

Where N = the number of students in the whole target population

n = the number of students in the whole sample for the target population

N_i = the number of students in the ith stratum of that population

n_i = the weighted number of "students" in the i^{th} stratum of the sample.

${}_1n_i$ = the weighted number of "students" in the first subsample.

[2] A full description of the weighting procedures used is given on pages 213 and 214 in Volume I of Husén *et al.*, 1967.

The calculations of means, standard deviations and correlations had to be carried out in terms of weighted N's.[3]

Standard Errors

Peaker in Husén *et al.*, 1967 (Volume I, Chapter 9, p. 154 *et seq.*) has explained in detail the calculations of both the simple random sampling (s.r.s.) standard errors and the complex standard errors (c.s.e.) of sampling.

Suffice it here to give Table 2.1, listing, for Populations 1a, 1b, 3a and 3b, a) factors by which the corresponding s.r.s. estimate should be multiplied to give the complex standard errors and b) complex standard errors for correlations.

The s.r.s formula for the standard error of a correlation coefficient is $(1-r^2)/\sqrt{n}$. The computer obtained the s.r.s. error for each population in each country first by comparing the average correlation coefficients obtained from four replicas (sub-samples) of a 54×54 correlation matrix with the four separate coefficients obtained and then averaging these for each matrix.

The s.r.s. formula for the standard error of a mean is, of course, $\sigma/\sqrt{N}$. To arrive at the c.s.e., the s.r.s. should be multiplied by the factor in the (a) columns in Table 2.1. It will be seen that the

[3] The following formulae for the weighted mean, standard deviation, and correlation were used:

Mean
$$ \bar{X} = \frac{\Sigma w_i X_i}{\Sigma w_i} $$

Standard deviation
$$ S = \sqrt{\frac{\Sigma((X-\bar{X})^2 w_i)}{\Sigma w_i - 1}} $$

Correlation
$$ r_{xy} = \frac{\Sigma((X_i-\bar{X})(Y_i-\bar{Y})w_i)}{\sqrt{\Sigma((X_i-\bar{X})^2 w_i)}\sqrt{\Sigma((Y_i-\bar{Y})^2 w_i)}} $$

where w_i = the weight for the ith student

X_i = the value of the X variable for the ith student

Y_i = the value of the Y variable for the ith student

Table 2.1 (*a*). *Factors* by which the corresponding s.r.s. estimate should be multiplied to give the complex standard errors and* (*b*) *complex standard errors for correlations.*

	Populations							
	1a		1b		3a		3b	
Country	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
Australia	1.7	.03	1.7	.03	**2.0**	.06	—	—
Belgium	1.7	.04	2.0	.04	1.6	.07	1.9	.06
England	1.7	.03	1.7	.03	1.3	.04	1.3	.03
Fed. Rep. of Germany	—	—	**3.3**	.05	1.3	.05	1.0	.04
Finland	1.7	.05	1.8	.05	1.3	.06	1.3	.06
France	2.1	.04	3.1	.05	1.1	.06	—	—
Israel	—	—	1.8	.03	**0.9**	.07	—	—
Japan	**1.4**	.03	**1.4**	.03	1.4	.05	**2.0**	.03
Netherlands	1.7	.08	1.9	.05	1.6	.07	—	—
Scotland	**2.9**	.04	3.1	.04	1.5	.04	1.8	.04
Sweden	2.3	.04	2.5	.04	1.6	.05	**0.9**	.05
U.S.A.	1.7	.02	1.7	.02	1.6	.04	1.8	.04
Mean	1.9	.04	2.1	.04	1.4	.06	1.5	.04

* In each of the factor columns (a) the highest and the lowest factor are in bold type.

average value of the ratios in Table 2.1 is 1.7, and that no ratios are very far from this value. Consequently, the rule of taking two (complex) standard errors as the confidence limits can be replaced by the rule of taking three s.r.s. standard errors.

Summary

This chapter has presented a short account of the history, structure and mode of operation of the first large scale international educational research project—namely that carried out by the IEA in the field of mathematics, from which the data for this study are drawn. It then proceeded to describe and define the target populations chosen for study and the sampling procedures used.

During the fifties there was a growing awareness on the part of some educators, and in particular educational research workers, of the need to establish evaluation techniques which would be valid cross-nationally. Groups of educational research workers from leading research centres in Europe and the United States joined together, and in 1959 undertook a small pilot project to test the feasibility and meaningfulness of carrying out cross-national educational research (see Foshay, 1962). Encouraged by their success, they embarked on a major research in the field of school mathematics education in 1962. They received financial support for their international costs from a grant from the United States Office of Education. National Research Centres were responsible for defraying the national research costs involved in the project. Research Centres from the following countries participated: Australia, Belgium, England, Federal Republic of Germany, Finland, France, Israel, Japan, the Netherlands, Scotland, Sweden and the United States. Each Research Centre had one member on the Council of IEA, whose task it was to agree on the overall policy of the research. Interim decisions were taken by a Standing Committee (elected from the Council), or by the Chairman and Technical Director. Since all persons involved had full-time commitments in their own countries, one full-time co-ordinator was appointed by IEA and placed at the UNESCO Institute for Education in Hamburg. Consultants were also employed and most of the work was undertaken by groups at Council Meetings, but some group work was also undertaken between meetings. Instructions were issued to National Centres in circular letters and special bulletins. There was a continuous two way communication between the research workers in the National Centres and the IEA Secretariat (Chairman, Technical Director and Co-ordinator). The analyses were carried out by computer at the University of Chicago Computation Center.

Four target populations were chosen which had to be sampled and tested by each participating Research Centre. These were

(a) all 13-year-olds (Population 1a)

(b) all students in the grade where most 13-year-olds were to be found within three months of the end of the school year (Population 1b)

(c) pre-university students studying mathematics as a major subject (Population 3a)

d) pre-university students *not* studying mathematics as a major subject. (Population 3b)

It was possible for Research Centres to test major terminal populations at points intermediate to the 13-year-old and preuniversity populations, but this was optional.

Probability sampling was used with the school as the sampling unit. In the United States, three stage sampling was used (community, school and students within schools), and in other countries two stage sampling (school and students within schools). Stratification was employed so as to reduce the intra-class correlation. The factors by which the corresponding simple random sampling (s.r.s.) estimates should be multiplied to give the complex standard errors are given, together with the complex standard errors for correlations.

CHAPTER 3

Instrument Construction, Data Collection and Processing

The aim of this chapter is to describe briefly the construction of the instruments. A very full description is given by Husén *et al.* (1967, *Volume I*) of the construction of the mathematics tests, questionnaires and occupational classification scheme, and the reader interested in further details is advised to refer to that publication.

Mathematics Tests

In order to formulate the general plan of the tests and the detailed specifications in terms of which they could be constructed, the following steps were taken, as described by Thorndike in Husén *et al.* 1967:

1. The research centre for each participating country was asked to recruit a committee of mathematics educators who would prepare a statement describing the content and objectives of mathematics education in that country.
2. These statements, so far as they were in fact prepared, were examined by a working committee of mathematicians and mathematics educators from several participating countries, and a topical outline was prepared covering the topics that appeared in the reports from the individual countries.
3. The outline was circulated to all participating countries, requesting judgements of the extent to which each topic was indeed covered in the mathematics instruction of the country.
4. On the basis of the responses, together with the judgement of the working committees, simple integral weights were assigned to indicate the importance and emphasis to be given to each topic.
5. In addition to preparing an outline of topics to be covered, attention was given to the types of intellectual processes to be covered.

6. The working committee developed plans relating to the number, length and types of test exercises to be included.

Each National Research Centre organized one or more committees to carry out a content analysis of what was taught in the various grades between Population 1b and the pre-university year, and in some cases the analysis was carried out by school type within a country. The work consisted mostly of an analysis of text books, examinations and teachers' statements. The documents produced by each National Centre were then sent to the International Mathematics Committee.

Two initial outlines were constructed, one for Level 1 (i.e. Populations 1a and 1b) and one for Level 3 (i.e. Populations 3a and 3b). Each outline contained about 40 different topics. A list of the topics for each level is given in Tables A2 and A3 in the Appendix. In each case, however, the objectives or categories of intellectual process were the same, namely:

A. Knowledge and information: definitions, notation, concepts;
B. Techniques and skills: solutions;
C. Translation of data into symbols or schema and vice versa;
D. Comprehension: capacity to analyse problems, to follow reasoning;
E. Inventiveness: reasoning creatively in mathematics.

In Tables A2 and A3 in the Appendix, the column headed *Objectives* indicates the categories of intellectual process that the working committee thought might be appropriately tested in connection with the various topics. The *Importance* column indicates the relative weight to be given in the final testing to each of the topics. (3 signifies great weight, 2 intermediate weight and 1 the least weight.)

Before preparing a pool of test exercises, the mathematics committee had to decide on the length, structure and format of the tests. Three to four hours of testing was accepted as a practical compromise between a comprehensive coverage and what represents a tolerable burden on the time of students and teachers. It was agreed, somewhat reluctantly, to keep the single problems brief. Much as one might like to explore the students' ability to work through an involved sequence of steps, or to develop a complex proof, this seemed not to be possible. Such a task would exhaust too large (and

too variable) a fraction of the limited time that was available. Thus, it was decided to limit the tasks to those that a student could be expected to deal with, if he could handle them at all, in not more than, and usually a good deal less than, five or ten minutes for each item.

The requirement of objectivity of scoring suggested the need to fall back on an all-or-none evaluation of a final product—the answer—and this was agreed, not without misgivings, since it was clearly recognised that the restriction placed real limitations on what could be appraised with the test. However, the decision seemed inevitable for an international study involving over a hundred thousand examinees. Furthermore, it was agreed to use mostly multiple-choice type items where the answer choices are supplied and the examinee chooses the best or correct answer. The committee recognised that there are many situations in which producing the response, rather than recognising it, is an essential part of the ability being tested. However, the practical necessity of speeding the scoring of the many papers called for machine scoring and for as extensive a use of multiple choice questions as seemed reasonable within the limits of effective measurement. In the end, 30 of the 174 items in the series required the examinee to write in his answer to a problem while 144 items were in multiple choice format. Using multiple choice items also had the advantage of allowing students to fill their answers in directly on to an IBM 1230 answer sheet which, with very little extra coding at the research centre, could be scored mechanically.

National research centres and members of the test committee supplied illustrative items for each of the topics in the test specifications. Using these items, and also items made available by the Educational Testing Service and by the University of Chicago Examiner's Office, a pool of some 640 items was assembled. Items were selected from this pool and 24 trial test forms were produced; the more elementary forms contained about 22 to 25 items and the more advanced forms 10 to 16 items. Each form was of such a length that it could be easily completed within 45 to 60 minutes. Two anchor items were included in all tests.

The trial test forms were then circulated to National Research Centres, and it was at this point that, as a result of criticism from England, additional trial forms were prepared. Finally, there were twenty-eight trial forms consisting of 497 items. The objective in preparing the trial forms was to make them inclusive, so that information might be obtained on a wide range of topics and formats.

Each trial form was then translated into the various languages, checked, and pre-tested on judgement samples of about 100 to 150 students in each country. Each test was pre-tested in at least three countries; the assignments were rotated so that different combinations of countries took each of the tryout forms. In each country eight or ten forms were pre-tested. According to the level of the test, it was tried out at the 13-year-old or pre-university level. In some, but not all, countries, appropriate tests were tried out at the 15/16-year-old level.

An item analysis was then carried out in the National Research Centres. Basically, this consisted of calculating the difficulty and discrimination indices estimated by Flanagan's procedure, for each item for a particular sample and reporting these back to the Test Editors. The results from all countries were then entered on to master tables.

The international test committee (Test Editors and Mathematics Educators) agreed that it was desirable to have some parts of the test common to the testing at the four different levels:

(a) 13-year-olds, and the grade group containing the largest fraction of 13-year-olds
(b) an intermediate age or grade group of roughly 15 or 16
(c) a group in the final year of secondary education, but not in a programme with mathematics as a major subject of study
(d) a group in the final year of secondary education *with* mathematics as a major subject of study

It was decided to organize the test in nine one hour units, each of which would be printed in a separate booklet and each of which would constitute a separate "test". The tests taken by each of the populations have already been given in Chapter 2 (see page 29). The items, 174 in all, were selected on the basis of their content validity to the test specifications and on their statistical attributes. In planning the content of the final tests, the editors attempted to maintain a balance between conventional content of mathematics and the newer topics that are being introduced in at least some of the participating countries.

Table 3.1 groups the items into topics in any one set of tests. In

Table 3.1. *Summary of content of tests for different populations.*

Topic	Popn. 1	Popn. 2	Popn. 3a	Popn. 3b
Basic arithmetic	13	3		3
Advanced arithmetic	18	7	3	9
Elementary algebra	12	6	1	5
Intermediate algebra	4	16	19	13
Euclidean geometry	13	17	5	13
Analytic geometry	1	4	8	5
Sets	4	3	4	4
Trigonometric and circular functions		1	3	3
Analysis			8	1
Calculus			9	
Probability			1	1
Logic		2	8	1
Affine geometry	3			

the final analysis, however, seventeen different sub-scores were calculated.

Estimates of the reliability of the total test and subscores were obtained for each population in each country, using the Kuder Richardson procedure of estimating reliability from item statistics and the standard deviation. Formula 20 was used.

Table 3.2 on page 44 gives the reliabilities for the Total Mathematics Score in each country for Populations 1a, 1b, 3a and 3b.

Although the analyses in this book are mostly concerned with Total Mathematics Score, it is of interest to comment on the various groupings of items. Firstly, they were classified, by the pooled judgement of several judges, into items calling for higher mental processes and those calling for lower mental processes. Lower mental process items are those which call for relatively routine application of previously learned techniques, while higher mental process items call for a greater amount of ingenuity and inventiveness in the attack upon novel or complex problems. A second subdivision of the items was into those that consisted of verbally formulated items, in contrast with those that involved primarily computation and solution of a problem expressed in numbers or symbols. A third sub-grouping of items consisted of those which were judged by the mathematics educators to represent the "new mathematics". Fourthly, items were grouped by content areas, i.e., arithmetic, algebra, geometry, etc.

Table 3.2. *Reliabilities of the total mathematics score for populations 1a, 1b, 3a and 3b in each country.*

Country	1a	1b	3a	3b
Australia	.913	.882	.867	—
Belgium	.929	.913	.906	.836
England	.951	.958	.923	.895
Fed. Rep. of Germany	—	.897	.848	.800
Finland	.888	.901	.865	.844
France	.929	.927	.913	—
Israel	—	.917	.817	—
Japan	.941	.941	.925	.926
Netherlands	.948	.915	.794	—
Scotland	.933	.940	.861	.844
Sweden	.869	.869	.897	.732
U.S.A.	.909	.906	.915	.844

Some statistical evidence was gathered on the validity of the IEA tests in England by comparing "O" and "A" Level students' performance on the IEA tests with their performance two or three months later in their "O" and "A" Level examinations. The average correlation was 0.65 for "O" Level and slightly higher for "A" Level, which indicates that there is substantial overlapping, but that it is far from complete. However, in the absence of information on the reliability of the G.C.E., it is not possible to state how nearly the IEA tests and the G.C.E. are measuring the same achievements.

Questionnaires

It was decided to collect information about as many relevant variables as possible that were likely to affect the mathematics performance of the students in the various countries. Among the most obvious factors are home, school and the structure of the educational system. The information about these environmental fields was collected from four main sources: the student, the mathematics teacher, the school principal and an expert on the educational system of each country. Accordingly, there were four types of questionnaires: a Student Questionnaire (ST 1 and 2), a Teacher Questionnaire (TCH 1), a School Questionnaire (SCH 1), and a National Case Study Questionnaire.

The data for variables on the students' background and schooling, collected by means of the Student Questionnaire, concerned such information as grade, sex, age, size of mathematics class, amount of mathematics instruction and homework, father's and mother's occupation[1] and education, aspirations and expectations for further mathematics, further schooling and occupation, best and least liked subjects, examinations taken and extra-curricular mathematics activities. The information requested from teachers concerned mainly teacher certification both in subject matter and professional training, teaching experience, recent in-service training, experience in "new mathematics" and teacher freedom. The information on school characteristics collected concerned school enrolment, number of male and female full-time teachers, number of trained mathematics teachers, type of school, the amount of educational expenditure, age range of students in school and school finance. The National Case Study Questionnaire[1] attempted to collect both quantitative and qualitative data concerning the students in full-time schooling according to school type, selection processes, compulsory schooling, economic data to determine the degree of economic, industrial and technological development and sociological data to determine the role of women in society. This latter questionnaire was completed by one person in each country who not only knew his own system well, but also had a good knowledge of other systems of education.

Only the Student Questionnaires were pre-tested. They were administered (at the same time as the mathematics trial forms were administered) to judgement samples of between 100 and 150 students in each country at both the 13-year-old level (ST 1) and the pre-university level (ST 2). Few modifications proved necessary. The Teacher, School and Case Study Questionnaires were not pre-tested but subjected to comments from experts in the field of questionnaire construction. Research Centres could, if they wished, add extra questions to the questionnaires for the purposes of a national survey.

It was, in some cases, necessary to adapt and modify certain ques-

[1] The construction of an occupational scheme is discussed in detail in Husén *et al.* (1966, Volume I, Chapter 8). Paternal occupation was chosen as the main indicator of family status. Nine categories of occupation were arrived at and agricultural occupations were given special categories within the nine. The difficulties involved in arriving at a classification scheme which is also a scale in all countries were formidable, but it was achieved in a limited way.

tions to national conditions so that a question was comprehensible to those answering it, or so that the information collected was comparable and thus more accurate than a mere translation of the international question; similarly, the source of information varied from country to country for some questions. Thus, for example, in some countries, the head teacher was able to give the data on teachers' salaries, but in other countries this information had to be collected from central records. Examples of the different ways in which the question concerning the extent to which ability grouping was practised within schools are given in Chapter 6.

The coding and punching schemes for the international questionnaires were drawn up by an international committee and these appear in Husén *et al.* (1967) as an Appendix to Volume I. The establishment of international codes was an extremely difficult exercise; the establishment, for example, of one common code into which all school types from all countries could be fitted proved much more difficult than expected, and much discussion and correspondence was required before all were satisfied with and understood the international codes. It should also be pointed out that a Student Opinionnaire was constructed, consisting of two environmental description instruments and five attitude scales, but since none of the data from the Opinionnaire are used in this book, its construction has not been described here.

Data Collection

Administration

It was extremely important to ensure that as far as possible uniform methods of procedure were employed in the testing programme in all countries, and also that very strict standardised procedures were used at the coding and punching stage. In order that this should be the case, a small committee prepared three manuals for National Centres' use. Manual 1 was designed to provide an adequate guide to National Centres concerning all the main procedures to be taken. It included a list of decisions to be made by National Centres, as well as suggestions for sub-sampling within schools and translating and printing the instruments; explanations of particular questions and their codes were also given, as well as instructions for sending all materials to the computing centre. The object was to indicate various methods of procedure to the National Centres in the field work,

and a uniform method of procedure at the coding and punching stage.

Manual 2 was a manual designed for the person responsible for the overall testing programme *within* any one school. The National Centre could decide whether or not it wished to use this in its original or modified form. This manual included a general account of what the project was, the timetable for testing (which varied from country to country), instructions concerning the receiving and storage of testing materials and preparation for the testing sessions, instructions concerning the lay-out of the testing room and the number of invigilators (proctors) required and the briefing of the test administrators and instructions concerning the return of all materials to the National Centre.

Manual 3 (which, again, could be used by the National Centre if so desired) was for test administrators and was the normal type of manual of instructions for test administration. If a National Centre desired to use Manuals 2 and 3 in a modified form, their proposed changes had first of all to be confirmed with the Technical Director.

The total testing programme comprised one and a half days' testing; this imposed a burden on a school, and for those schools where students at different levels were being tested, this burden was considerable. In some of the countries no national survey of this kind had previously been undertaken. This was, therefore, a first experience in large-scale test administration for some National Centres and for the schools, teachers and students in those countries. Difficulties were, of course, experienced, but the results of the experience were encouraging in that few data were lost because of difficulties met in the administration process. It was interesting to note that some National Centres, in whose countries answer sheets had not previously been used, decided to use them. The operation turned out successfully and no difficulties were experienced; the instructions given in Manual 3 on how to fill in the answer sheets appeared to be clear and comprehensive. Apart from the manuals, further instructions were sent out in circular letters, and the main points were every so often summarised in bulletins.

In most cases, the testing in the classrooms was carried out by teachers, but there were exceptions; for example, in Belgium members of the psycho-socio-medicaux centres who are trained in test administration were employed. In Finland, members of the Department of Educational Research of the University of Jyväskylä each took responsibility for the schools in a particular area. The department supplied them with cars, and they completed the testing programme within two weeks.

Data Recording

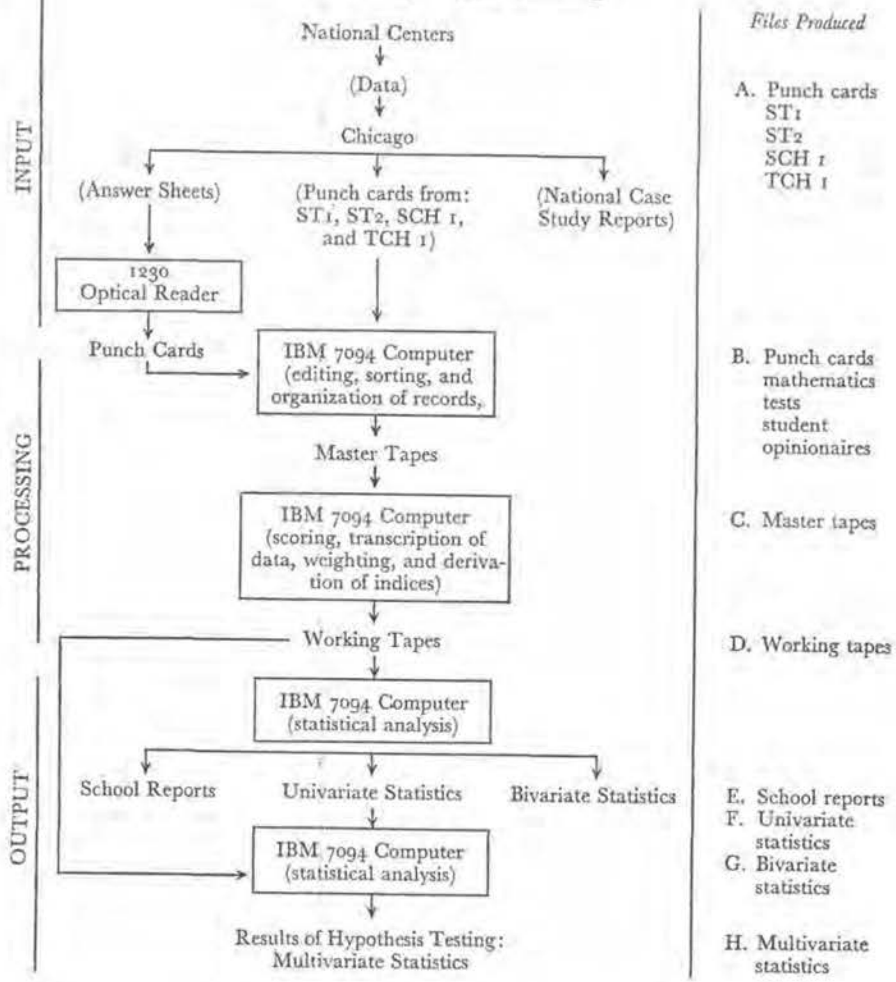
The material from each school was sent to the National Centre. For the laborious and painstaking work of recording the data from the questionnaires on to punch cards or on to special answer sheets designed for the IBM 1230 machine (which then produced a punched card for each answer sheet), each National Centre either employed some of its own staff or hired special staff to do the coding, all of whom worked under the supervision of the person responsible for the IEA project in each centre. Certain questions were asked in different ways in different countries, and it was, as has already been pointed out, of paramount importance that the information given in response to each question was recorded in a standardised way from country to country. For this reason, the responses to as many questions as possible were pre-coded. Where postcoding was required, the columns and ranges on columns (i.e., number of punch positions) were specified.

To ensure standardised recording of data, certain check procedures were set up, which involved National Centres sending their own coding and punching scheme for checking to the IEA secretariat. After this had been approved and when coding and punching had begun at the National Centre, the first twenty punch cards of each type of questionnaire, plus copies of the questionnaires, were sent to Chicago for checking.

National Centres were informed of any errors picked up in these two checks and were asked to correct them before coding and punching of all the questionnaires proceeded.

After all the answer sheets and punch cards were ready, they were despatched to Chicago where all data were entered on to magnetic tape at the University of Chicago Computation Center. When it is realised that in all twelve countries together, 132,775 students from 5348 schools were tested, and that questionnaires were filled in by 13,364 teachers and 5348 headteachers, it will be appreciated that the amount of time required to record these data at National Centres was enormous.

Flow Chart for Data Handling



Data Processing

Although the first data arrived in Chicago in September, 1964, programming had already been underway for a good nine months. The main programmes to be written (apart from programmes for specific hypothesis testing) were the editing, sorting and filing programme, and the programme for compiling the working tapes from the master tape. On the arrival of the Answer Sheets, there was a considerable delay, since it turned out that about one-fifth of all the Answer

Sheets had to have their responses "re-blackened", and a certain number of Answer Sheets had to be completely recopied, since their edges had been damaged in transit.

The data (approximately fifty million pieces) were entered on to the master tape in their raw form (i.e. every response to every item by every individual—student, teacher, head teacher and national case study expert—at every level in every country). Four edited working tapes were compiled, one for each population. All mathematics scores were weighted (see Chapter 2) and corrected for guessing on the working tape, and mathematics sub-scores and various derived indices have been produced. Analyses were then carried out in two stages: first, univariate and bivariate statistics were produced for each population in each country; second, specific hypotheses were tested, as well as a multiple regression analysis being run. The computer used throughout was an IBM 7094. The flow-chart on page 49 may be useful in understanding the total processing system.

Summary

The steps taken in the construction of the mathematics tests were:

(a) content analysis of mathematics courses and statement of objectives of mathematics

(b) preliminary outline of topics and objectives drawn up as test blue-print

(c) topics weighted and test blue-print produced

(d) four hundred and ninety-seven trial items formed into 28 pre-test forms

(e) fourteen pre-tests tried out at 13-year-old level and fourteen at pre-university level on judgement samples of approximately 150–200 students at each level. Each test was tried out in at least four countries. In some countries some tests were also administered to 15/16 year-olds.

(f) item analysis

(g) ten final tests (174 items) constructed such that one test was common to at least two different populations. A maximum of 17 different sub-scores could be computed.

(h) evidence of the concurrent validity of the IEA tests in England was collected for two populations. The average correlation was about .65.

Background information was collected on students by means of a student questionnaire, one version being administered to 13-year-olds (ST 1) and another to the pre-university students (ST 2). These were pre-tested on judgement samples of approximately 100 students in seven countries. Very few changes were required. Background information on the students' teachers and schools was collected by means of a teacher questionnaire (TCH 1) and a school questionnaire (SCH 1). Neither of these was formally pre-tested, but each was worked out by experienced questionnaire constructors. All questions and codes were found to work satisfactorily. Some difficulty was experienced in the establishment of international codes, but it was found that the "common moulds" eventually proved appropriate. Data to provide a contextual background for the findings of the research in terms of the school system and societal and economic factors etc. were collected by means of a National Case Study Questionnaire completed by a national comparative educationist.

Three different manuals were produced for use by National Centres, school testing organisers and actual testers, so as to ensure standardisation of procedure throughout all the full testing programme and coding and punching stages. In most cases, the actual testing was carried out by teachers, but in some cases was carried out by trained testers or by students of psychology or education.

All responses to the mathematics items were recorded on specially prepared IBM 1230 answer sheets. Responses to questionnaire items mostly pre-coded, but some required post-coding) were punched on punch cards at the National Centre, but only after a series of checks had been carried out on the punching of the first twenty of each type of questionnaires. Answer Sheets and punch cards were then sent to the University of Chicago Computation Center and there all responses were entered on to a master tape. Working tapes were compiled, involving the weighting of scores and the derivation of sub-scores and special indices. Analyses were then carried out in two stages—the production of univariate and bivariate statistics and the testing of specific hypotheses.

CHAPTER 4

The Investigative Situation

The problems examined in the present study will be viewed against the background of the school organization of the countries included. The aim of this chapter, therefore, is, firstly, to describe briefly the structure of the educational systems participating in the study, and secondly, to describe in some detail various aspects of the systems relevant to the features of school organization taken up in Chapters 5 to 7.

Before noting the differences between the structures of the systems, it is worth mentioning several features which they obviously have in common. All have universal primary education. All are high income, technologically and industrially developed nations when compared with the world as a whole. All have a tradition of education.

Apart from the differences in the structures, it is necessary to state that the geographical and cultural contexts in which these structures are to be found vary widely. No evidence which is used in this study is concerned with national socio-cultural differences, and measures of such cultural differences, will, therefore, not be dealt with here. What then are the major differences in the school structures? The first difference concerns the age of entry to school. This varies from five years of age in England and Scotland (which differ in their overall structures as can be seen from Figures 4.3 and 4.8), and seven years of age in Finland and Sweden. Since in Chapter 7, the problem is taken up of the association between mandatory age of entry to school and mathematics scores at age 13, it should be pointed out that within limits, whereas children entering school at five in England and Scotland are gradually led towards the formal type of lesson, in other countries there tends to be a formal type of schooling imposed fairly quickly. Furthermore, there is considerable variation between countries in the proportion of an age group which attends nursery school or kindergarten (cf. Chapter 7).

The second major difference is that some systems practise inter-

school grouping, where others do not. The former systems select a percentage of an age group (ranging from 15 to 25%) at a certain age out of the main school into a selective-academic school. The age of selection ranges from ten in the Federal Republic of Germany, to twelve in Scotland; the mode of selection also varies from ability and achievement testing plus interviews (for some) in England to teachers' judgements alone in other countries. There is evidence to indicate that these forms of selection are associated with social factors even when "objective" selection instruments are used (Undeutsch, 1960; Halsey, 1961; Douglas, 1964; Husén, 1966). The latter systems have no different types of institutions during compulsory schooling and all children, irrespective of social origin or academic ability, proceed through the school without being separated from their peers. It is only towards the end of the compulsory term of schooling that some degree of differentiation of programme is allowed.

Figs. 4.1–4.10. *National Systems of Education*

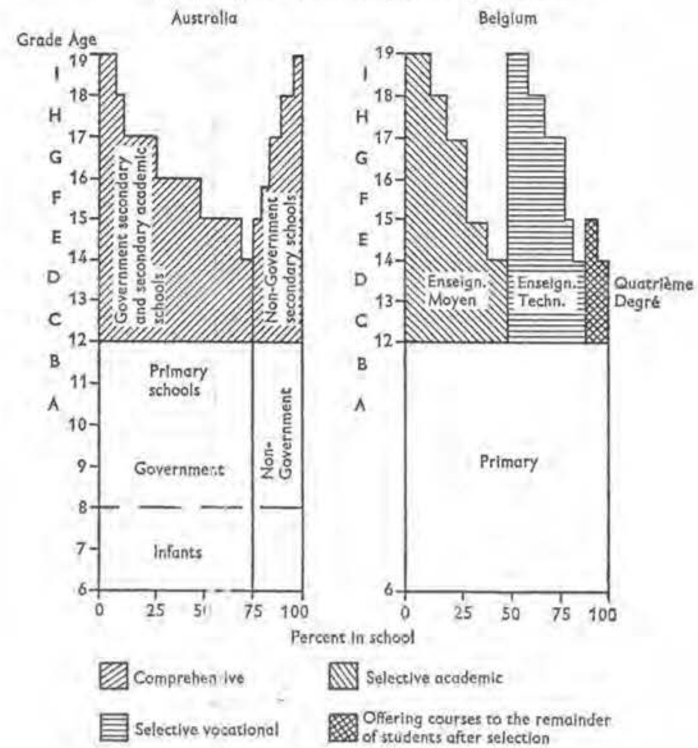


Fig. 4.1 *Fig. 4.2*

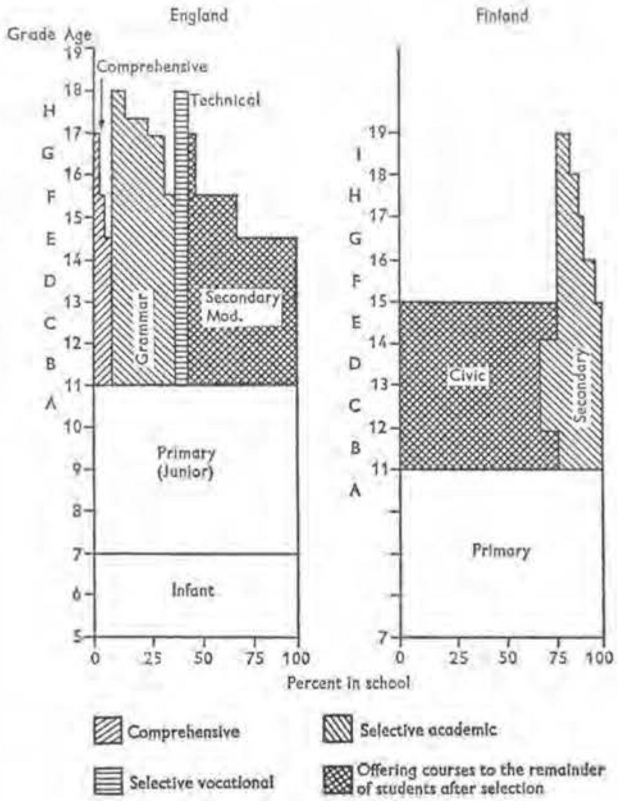


Fig. 4.3 *Fig. 4.4*

Structures

Before proceeding to comment in more detail on some of the differences, a set of figures is presented indicating the proportions of children in full-time education and how these are distributed among major school types within countries.[1] The figures are based on data collected in the National Case Study Questionnaire as well as from the Unesco World Survey (1961), where this was relevant. Although the names of the types of school have been given, the school types are also designated as belonging to one of four categories: comprehensive, selective-academic, selective-vocational or remainder. The

[1] Similar discussions on this point are to be found in Postlethwaite, 1965 and Husén *et al.*, 1967.

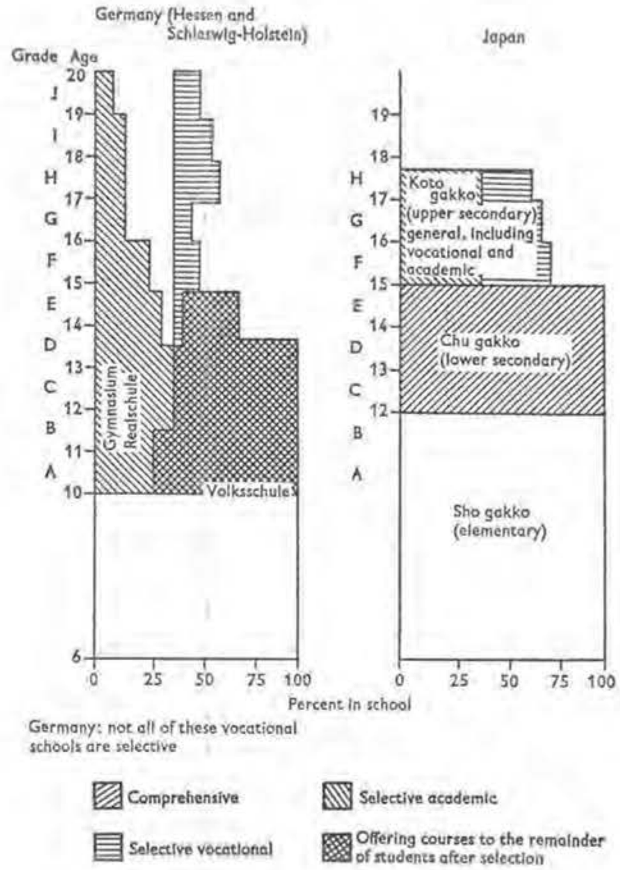


Fig. 4.5 *Fig. 4.6*

first three categories are self-explanatory; by remainder is meant the type of school which those students attend who are not selected out in a selective system (e.g. *Secondary Modern School* in England, *Volksschule* in the Federal Republic of Germany, etc.) The proportions still in school are proportions of an age group. The grades in which most of an age group are to be found are given by the side of the age group. Grade D is Population 1b in each country (see Table 4.1).

In connection with the figures on pages 53–57 and also with Table 4.3, it should be mentioned that a) in Australia at the age of eighteen there is a large decline in the proportion of an age group in

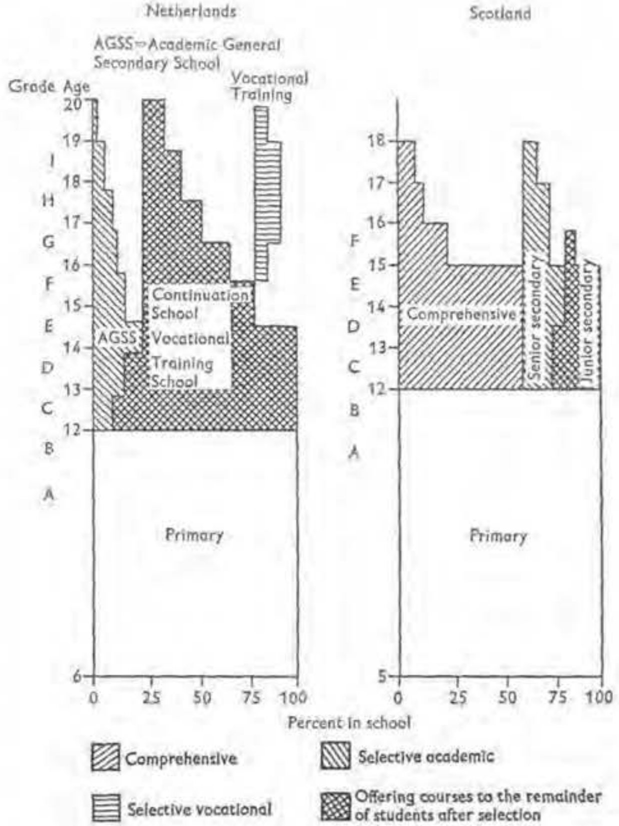


Fig. 4.7 *Fig. 4.8*

schools and that this is partly attributable to the low age of entry to institutions of higher education; b) in Germany not all of the vocational schools are selective and c) by 1970 in Sweden, all children up to the age of 16 will be in comprehensive schools.

From the figures it is possible to see the different ages of starting school, the point at which selection takes place (if it does at all) and the approximate percentage of an age group remaining in school through the various grades and in various school types to the end of secondary schooling. Although more detailed comment is made in Chapter 7 on the mandatory age of starting school, it would be useful to provide a separate table indicating the median age of entry to school, the mandatory age at which compulsory schooling ends and the average age of students three months before the end of the pre-

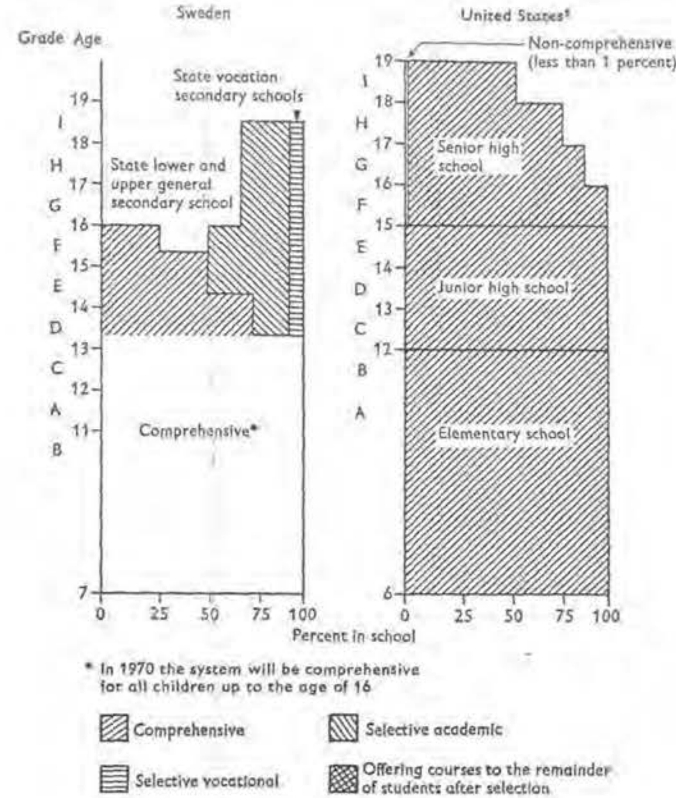


Fig. 4.9 *Fig. 4.10*

university year. The source of the first two pieces of information is the National Case Study Questionnaire, whereas the last piece of information comes from the Student Questionnaire. The data are presented in Table 4.2.

It must be remembered that the degree of pre-schooling (nursery school, kindergarten, etc.) varies from country to country—see Chapter 7. Furthermore, whereas in most European systems there is only one entry point to school each year, in England and Scotland there are two or three. There is evidence from England (Douglas, 1964; Pidgeon, 1965) that the multiple points of entry, together with other factors of school organization, affect the size of the standard

[a] It should be noted that although the most frequent form of school organization has been shown here, namely the 6-3-3, other forms do exist: 6-2-4, 8-4, 5-3-4 and 5-4-3.

Table 4.1. *1b populations—designation of grades.*

Australia	1st Form—in New South Wales, Queensland, South and Western Australia 2nd Form—in Victoria and Tasmania
Belgium	5e (2e A3 in *Enseignement Technique*)
England	3rd Form
Fed. Rep. of Germany	7. *Klasse* (*Schulleistungsjahr*)
Finland	7 in primary school 1 in civic school 3 in secondary academic school
France	5e (C.S.E. in *école primaire*)
Israel	*Khet* (8th Grade of elementary school)
Japan	*Ni-nen* 2nd Grade
Netherlands	6e in primary schools 1e in other schools
Scotland	2nd year of secondary course (S2)
Sweden	*Årskurs* 7
U.S.A.	8th Grade

deviation of an age group on scores obtained on achievement and ability tests administered at, for example, the age of 8 or 11.

Similarly, although all countries stipulate a minimum age which students must attain before leaving school, there is only one point of exit per year in some countries and two or three in others. It should be noted that there is a general movement in most countries to increase the statutory leaving age and that by 1970 few of the countries which have participated in this present study will have a statutory leaving age below 16. The differences between the average ages in Populations 3a and 3b are of interest, but an explanation other than that of sampling fluctuation is difficult to find.

The amount of inter-school differentiation being practised in the various systems can be seen from the Figures 4.1 to 4.10. It should be strongly emphasised that those schools termed comprehensive include a variety of schools, ranging from those where all children from an area attend, but are strictly divided into ability groups within the school (e.g. some Comprehensive Schools in Scotland), to those where all children are in heterogeneous groups at least to the age of 13 (e.g. Swedish Comprehensive Schools). The average amount of ability grouping practised within schools in each of the participating countries is given in Chapter 6.

Table 4.2. *School: Median age of entry, mandatory minimum age of leaving and average age of completing pre-university year.*

	Median age of entry	Mandatory minimum age of leaving	Average age of completing pre-university year	
			3a	3b
Australia	5 yrs 7 mo.	14–16 years	17 yrs 2 mo.	—
Belgium	6 yrs 2 mo.	14 years	18 yrs 1 mo.	18 yrs 0 mo.
England	5 yrs 2 mo.	15 years	17 yrs 11 mo.	17 yrs 11 mo.
Fed. Rep. of Germany	6 yrs 9 mo.	15 years full time 18 years part time	19 yrs 10 mo.	19 yrs 9 mo.
Finland	6 yrs 8 mo.	15 years	19 yrs 1 mo.	19 yrs 2 mo.
France	6 yrs 3 mo.	16 years	18 yrs 7 mo.	18 yrs 9 mo.
Israel	6 yrs 3 mo.	14 years	18 yrs 2 mo.	—
Japan	6 yrs 6 mo.	14 years	17 yrs 8 mo.	17 yrs 8 mo.
Netherlands	6 yrs 5 mo.	14 years	18 yrs 2 mo.	18 yrs 7 mo.
Scotland	5 yrs 2 mo.	15 years	17 yrs 6 mo.	17 yrs 1 mo.
Sweden	7 yrs 1 mo.	16 years*	19 yrs 7 mo.	19 yrs 7 mo.
U.S.A.	5 yrs 8 mo.	16 years (Some states approximately 18 yrs)	17 yrs 9 mo.	17 yrs 10 mo.

* According to 1962 Education Act.

Attrition Rate

Although it is possible to gain an approximate idea of the attrition rate from Figures 4.1–4.10, it would be useful to examine the various attrition rates in more detail. In Chapter 5, the mathematical "yields" (or "outputs") of several systems are examined, but these refer only to those still in school. Thus, for example, although it is interesting to compare the "yields" of those in school, this approach has limitations, since it would obviously be of interest to know the "yield" of those who have "dropped out" of school. This was not done in this study, but it is important to be aware of the varying proportions of students "dropping out" in the participating countries. In systems where students progress through the school more or less in age groups (e.g. England, Japan and Scotland), it is easy to see how many have participated both how long and how far in the systems. Unfortunately, in systems where grade repetition is frequent, or where advanced placement is common, or again where students may have begun school earlier than the mandatory age of entry to school, it is difficult, after looking at either the age or grade drop-

Table 4.3. *Proportion of boys and girls of the total age group in school and by grade.*

Country	Sex	Age								Grade					
		13	14	15	16	17	18	19	20	D	E	F	G	H	I
Australia	B	100	92.1	69.9	40.1	19.5	7.2	—	—	51.4	51.9	52.4	56.1	59.3	
	G	100	90.0	61.6	31.4	11.8	2.5	—	—	48.6	48.1	47.6	44.9	40.2	
Belgium	B	94.4	84.7	67.5	67.4	44.3	27.1	15.1	9.8	49.7	51.8	54.5	56.5	59.5	
	G	97.1	80.7	63.1	56.0	33.3	17.5	7.4	5.0	50.3	48.2	45.5	43.5	40.5	
England	B	100	100	43.4	23.7	13.5	5.4	0.6		51.1	54.1	52.5	53.8	57.3	
	G	100	100	41.0	21.1	10.5	2.8	0.2		48.9	48.9	47.5	46.2	42.7	
Federal Rep. of Germany	B	100	83.5	56.2	31.1	14.9*	15.7*	14.2*	7.0*	51.6	51.1	52.1	49.2	57.5	61.8
						18.6**	27.0**	24.5**	17.0**						
	G	100	83.5	55.1	29.6	11.0*	9.1*	7.1*	2.3*	48.4	48.9	47.9	50.8	42.5	38.2
						14.6**	14.0**	11.4**	5.8**						
Finland	B	99.6	98.0	40.2	27.0	20.0	14.2	9.3	3.8	48.8	49.0	43.8	43.8	43.8	43.8
	G	99.8	98.8	45.9	35.0	27.1	19.4	10.3	3.9	51.2	51.0	56.2	56.2	56.2	56.2
France	B	Not available								50.2	45.3	45.1	47.4	52.9	
	G	Not available								49.8	54.7	54.9	52.6	47.1	
Israel	B	Not available								50.8	50.1	50.9	50.9	50.5	
	G	Not available								49.2	49.9	49.1	49.1	49.5	
Japan	B	99.8	99.8	64.9	60.1	56.3				51.0	51.0	51.7	51.2	50.8	
	G	99.9	99.9	63.2	60.7	56.8				49.0	49.0	48.3	48.8	49.2	
Netherlands	B	100	86.8	72.6	60.4	47.0	32.7	21.1	13.9	Not available					
	G	99.1	78.9	50.4	30.4	19.6	11.8	8.0	4.9	Not available					
Scotland	B	Not available								Not available					
	G	Not available								Not available					
Sweden	B	95.6	79.7	55.9	45.1	34.6	28.3	16.2	11.0	51.5	49.5	47.1	49.3	51.8	59.2
	G	96.1	83.7	59.9	46.3	34.4	28.0	17.8	11.2	48.5	50.5	52.9	50.7	48.2	40.8
U.S.A.	B	96.9	95.4	93.0	86.5	74.8				50.8	51.0	50.8	50.2	50.7	
	G	97.0	95.3	92.6	86.0	74.3				49.2	49.0	49.2	49.8	49.3	

* Academic   ** Vocational

out figures, to have more than a general picture of how many students participate how far. For example, in Germany, students begin leaving school after the age of 13, but Grade E (the post 13-year-old grade) has an estimated hundred percent of an age group still in school. This is due to early starting school and to advanced placement.

Table 4.3 gives both the age group "drop out" by sex, and at the same time the proportion of boys and girls in each grade for each of the countries in the study, except for Israel; there are no figures made publicly available for Israel. The figures were those which were the most recently available in 1964 and in all cases are post 1960. Grade D is the grade in which most 13-year-olds were to be found when the testing took place (i.e. Population 1b). For Germany, the figures for the last year in school for both the secondary academic schools and the vocational schools are given, although it is only the secondary academic schools which are considered in this study.

Many more details are given on the age and grade drop-outs in each of the participating countries in Postlethwaite (1965), but Table 4.3 gives sufficient information for it to be seen that in the United States and Japan (where large numbers continue through to the end of the pre-university year) approximately equal proportions of boys and girls drop out, whereas in all other countries (with the exception of Finland) proportionally more girls than boys drop out. It is also interesting to note that some countries have succeeded in persuading fairly high proportions of an age group to elect to continue in school past the statutory age of leaving; of particular note is the regularity of the drop-out in Belgium and Sweden.

Specialization

The average number of subjects studied in each grade in secondary schooling varies from country to country. In England, for example, it is the custom for students to study up to nine or ten subjects (or sometimes more) until the age of 15 or 16, when they either leave school or take the first major national examination, the G.C.E. "O" level examination; thereafter, they tend to study only three or four subjects. In other countries, such as Belgium, Germany and Finland, as many as nine or ten subjects are studied right through to the pre-university year.

In Chapter 7, an analysis is carried out in which one classificatory variable is the average number of subjects studied by pre-university students in each of the participating countries. However, it is also of interest to note the average number of subjects studied in the grades preceding the pre-university year.

Table 4.4 sets out the average number of subjects studied in the pre-university year and the four preceding years in the secondary academic schools or programmes. The countries are ordered according to the average number of subjects studied in the pre-university year.

The figures for the United States may appear surprising, but it must be remembered that because of the system of credit points,

Table 4.4. *Average number of subjects studied in last five grades of secondary academic schooling**.

	X-4	X-3	X-2	X-1	Pre-university grade (X)
Belgium	9+	9+	9+	9+	9+
France	9+	9+	9+	9+	9+
Netherlands	9+	9+	9+	9+	9+
Japan	9+	9+	9+	9+	9+
Finland	9	9	9	9	9
Fed. Rep. of Germany	9	9	9	9	9
Sweden	9	9	9	9	9
Israel	—	9+	9+	8	8
Australia	8	8	8	7	6
Scotland	8	7	6	5	4
U.S.A.	5	4	4	4	4
England	9+	8	8	3	3

* Source for these data was question 14 on the National Case Study Questionnaire.

compulsives and electives in the Senior High School, it is unlikely that the subjects noted here ("solids") will be the same from year to year. In general, the figures in Table 4.4 indicate that, from the countries participating in this study, England and Scotland have adopted specialisation, whereas the other countries have continued general education, with the exception of Australia and the United States, which are half-way between.

Summary

The results of the present study must be viewed against the background of the school organisation of the participating countries. Of particular interest for the problems investigated here are the ways in which the students progress through the system, the points at which selection takes place, and the percentages of students in the different forms of schools, in particular, comprehensive, selective academic, selective vocational and other school types. (Figures and tables indicate these features for each of the systems.) In general, both the United States and Japan can be said to be retentive in that they have well over half of a year group continuing through to the pre-university year. Sweden, however, has recently changed from the traditional European dualistic pattern of education to the comprehensive, but in Scotland, although it has a high proportion of so-called comprehensive schools, the system of education is still basically dualistic, since the dualistic pattern is preserved within the comprehensive school through the practice of educational differentiation. Similarly, it must be remembered that in England many of the "comprehensive" schools do not contain students from the full distribution of ability within the areas, as often the top ten to twenty percent in terms of ability are attending a local *grammar school*. Although in Germany there are some students who attend *Fachschulen, Berufsschulen* or *Ingenieurschulen* full time in the last year of secondary education, these have not been considered in this study.

The attrition (drop-out) rate after compulsory schooling tends to be very high in selective countries, but it is interesting to note how regular the drop-out is in both Belgium and Sweden. The age composition of the pre-university year also varies greatly from system to system. In the United States, Scotland and Japan it is low, and in Sweden, Germany and Finland it is high.

A further important factor to be taken into account when comparing systems is the amount of specialisation in the last year of secondary education. England and Scotland are highly specialised (three or four subjects), whereas in other countries students study at least six subjects and usually nine or more.

CHAPTER 5

Retentivity

As was seen in Chapter 4, the attrition rate and amount of attrition differs considerably among the countries represented in this project. In general, the USA and Japan have highly retentive systems of education in the sense that a high proportion of each year group continues through to the end of secondary education. In Europe, on the other hand, there is a much smaller proportion of a year group proceeding to the pre-university year. The different proportions are connected with the different philosophies of comprehensive and selective school systems as well as reflecting differing socio-economic structures between the countries. Secondary education in most European countries has been characterised, until recently, by the selection and transfer of "more able" pupils into separate types of academic school while the rest of the pupils have remained in schools initially designed to provide a basic education for the majority of children (e.g. *elementary scool, Volksschule, école primaire*).

The academic secondary school, with a long tradition going back to the medieval Latin school, has tended to recruit (select) the bulk of its pupils from the higher socio-economic strata. On the other hand, the development of public education in most parts of the United States has not been markedly affected by traditional practices, with the result that the eight year elementary schools were not regarded primarily as a preparation for secondary schooling, but as self-contained establishments capable of extending their provision to satisfy the educational needs of the community. Thus, in the European school systems, there developed the practice of selecting an élite to go through to the pre-university year, whereas in the more comprehensive systems (e.g. U.S.A.) the type of system was such that there grew up a deliberate policy of encouraging as many pupils as possible to continue through to the pre-university year (cf. Husén, 1962).

However, many of the European countries are at present revising their policies. Economic growth and the recent rapid advances in science and technology have created the need for a more prolonged period of general education for all young people and not just for the most able minority, with the result that successive increases in the duration of compulsory schooling have been made in most European countries. Furthermore, the need for more skilled and better informed manpower has also resulted in a substantial increase, in most countries, in the numbers of young people choosing to continue their education beyond the statutory school-leaving age. In Sweden, for example, in 1950 only ten percent of seventeen-year-olds proceeded to *gymnasiet*, while by 1964 the proportion had risen to twenty-eight percent (Yates, 1966). By 1970, it is estimated that nearly 30 percent will wish to enrol in *gymnasiet* (Dahllöf *et al.*, 1966). This increased proportion of a year group continuing to the end of secondary education is often accompanied by a restructuring of the educational system itself, either by the introduction of a comprehensive system of education with no selection or by delaying selection into the academic secondary school.

In the Case Study Questionnaire, data were collected on the actual number of students in each year group still in full-time schooling, as well as the actual number of students in each grade group. The national statistics which were the sources of these data were, in general, available, depending on the country, for the years between 1960 and 1963. In every case, it was the most recently available statistics which were used. Furthermore, the heads of National Centres were asked to estimate for 1964, at the time of testing, (a) the percentage of an age group in school at the pre-university level and (b) the proportion who were specialising in mathematics (enrolled in the terminal Mathematics-Science programmes). The division into mathematics and non-mathematics students in the pre-university year has already been discussed in Chapter 2. It would seem that in some National Centres approximations were made to the nearest whole number, whereas in others, the proportion was calculated to the first decimal place. The actual figures supplied are used in this analysis.

These figures are given in Table 5.1 in which there are also given, in the fourth column, measures of the degree to which each country has adopted a comprehensive system of education. This has been

assessed by the percentage of students in the younger and complete age group (Population 1a) attending so-called "comprehensive" schools. This information was collected by means of the School Questionnaire (see Appendix II, Volume I of Husén *et al.*, 1967). A comprehensive school was described as offering appropriate courses for students of *all* ranges of ability.

From Table 5.1 it can be clearly seen that there is considerable variation among the countries in this study in the percentage of a year group continuing through to the pre-university year. Since it has been possible to measure the mathematics achievement of the pre-university students as well as the 13-year-old students[1] in the countries, it is worthwhile posing several questions concerning the amount of mathematical achievement of both the pre-university groups (in terms of the percentage of a year group still in school) and the 13-year-old group of students.

Table 5.1. *Indices of retentivity and comprehensive education.*

Country	Retentivity (percentages of age group) Total	3a	3b	Comprehensiveness (percentages of Pop. 1a)
Australia	23	14	9	70
Belgium	13	4	9	0
England	12	5	7	9
Fed. Rep. of Germany	11	4.7	6.5	0
Finland	14	7	7	0
France	11	5	—	0
Israel	—	7	—	96
Japan	57	8	49	100
Netherlands	8	5	3	0
Scotland	18	5.4	12.6	44
Sweden	23	16	7	64
U.S.A.	70	18	52	92

The rank correlations of the three indices of retentivity with the extent to which pupils are being educated in comprehensive schools are 0.89, 0.76, and 0.73 respectively.

1 For descriptions of the pre-university populations see pp. 237–239 of Vol. I of Husén *et al.* (1967). For description of the 13-year-old grade group see p. 29 in this book.

First of all, it is possible to examine whether there is a difference in the average score of students (in both of the two pre-university populations) in systems with different amounts of retentivity, i.e., *if more students are allowed through will this lower the average standard of performance?* Secondly, it is possible to examine the relative performances of the students by certain international standards by taking the number of students above the 95th international percentile and then discovering, for each nation, (a) what percentage this is of the students in full-time schooling and (b) what percentage these students are of a year group. This analysis will assist in an examination of the problem of *whether or not the standard of performance of the best students in the pre-university year deteriorates if a larger percentage of an age group goes through to the pre-university year.* Thirdly, it is possible to *examine the mathematics performance "yield"* of the target populations in the study. By "yield" is meant *how many* students are brought *how far* (in this case in terms of mathematics achievement as measured by the IEA tests), within the framework of full-time schooling in the educational system. This takes into account both the number of persons (in terms of the percentage of an age group reaching a particular level) and the level of achievement per person, and is therefore not simply a comparison of means between countries, irrespective of the differing percentages of an age group making up the population being compared. In this last case, it is also possible to compare increase in "yield" between the 13-year-old age group (where virtually one hundred percent of an age group are still in school) and the pre-university group of students. Thus there are three main problems, all of which are related to retentivity, which will be examined: Average performance, Fixed international standards performance and Yield.

In this connection it should be pointed out that there are differences on some major independent variables among the pre-university populations in the countries participating in this project. There is a wide variation in the socio-economic status composition of this group, ranging from a composition somewhat similar to the general population in the U.S.A., to a predominantly middle-class composition in Germany. A second major disparity is the mean age[2] which ranges from 17 years 2 months in Australia to 19 years 10 months in

[2] For a different analysis of age, retentivity and score see p. 68 *et seq.*, in Husén *et al.* Vol. II (1967).

the Federal Republic of Germany. A third variation lies in the average number of subjects studied in the pre-university year, ranging from three in England to nine or more in Belgium, France, Japan and the Netherlands. These discrepancies have been dealt with to some extent in Chapter 4 of this book and in much more detail in Chapter 2 of Volume I of the IEA publication (Husén *et al.*, 1967).

In the discussion of yield, Population 1b has been used rather than Population 1a, although the latter would have been better since it is a chronologically comparable group. However, four countries (Australia, France, Israel and the Netherlands) were lost at the pre-university level, since either they did not test Population 3b, or their sampling procedures were considered to be inadequate. If 1a had been chosen for the lower level rather than 1b, there would have only been seven countries left, since Germany did not test 1a. Hence, Population 1b was chosen.

Average Performance

The percentages of an age group still in school (circa 1964) in the two pre-university populations have been given in Table 5.1. The

Table 5.2. *Total mathematics score, means, standard deviations and N's for populations 3a and 3b.*

Country	Pre-university math-science programme Population 3a			Pre-university non math-science programme Population 3b		
	M	S.D.	N	M	S.D.	N
Australia	21.6	10.5	1089	—	—	—
Belgium	34.6	12.6	519	24.2	9.5	1004
England	35.2	12.6	967	21.4	10.0	1782
Fed. Rep. of Germany	28.8	9.8	649	27.7	7.6	643
Finland	25.3	9.6	369	22.5	8.3	399
France	33.4	10.8	222	—	—	—
Israel	36.4	8.6	146	—	—	—
Japan	31.4	14.8	818	25.3	14.3	4372
Netherlands	31.9	8.1	462	—	—	—
Scotland	25.5	10.4	1422	20.7	9.5	2123
Sweden	27.3	11.9	776	12.6	6.2	222
U.S.A.	13.8	12.6	1568	8.3	9.0	2042

means, standard deviations and N's for Populations 3a and 3b are given in Table 5.2.

The relation of mathematics score to the percentage of an age group in school by country is shown for Populations 3a and 3b in Figures 5.1 and 5.2 respectively. The rank correlations between the mean score and the percentages of an age group in school in each population are –.62 and –.36 for Populations 3a and 3b respectively. The decrease in mean score as the percentage of an age group retained in school increases is clearly discernible in both populations, giving weight to the contention that the greater the retentivity, the lower will be the average score of those retained. It might also be thought that the smaller the percentage of an age group retained,

Fig. 5.1. *Relation of Mean Mathematics Score to Percentage of Age Group in Population by Country*

(Population 3 a)

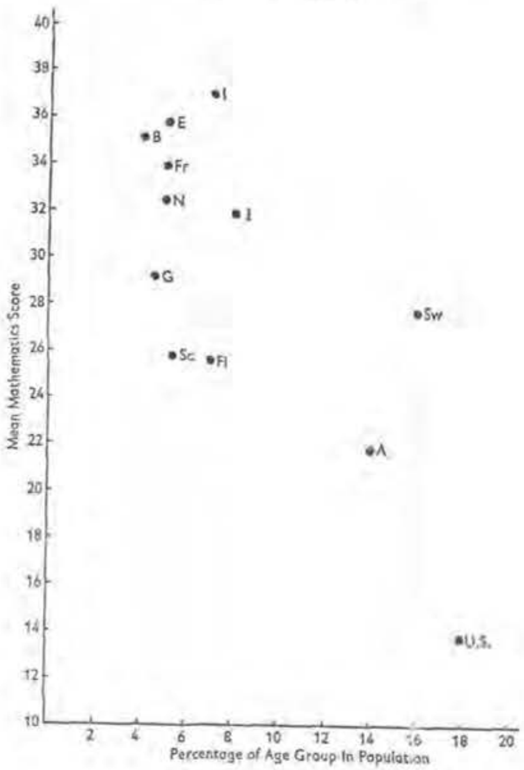
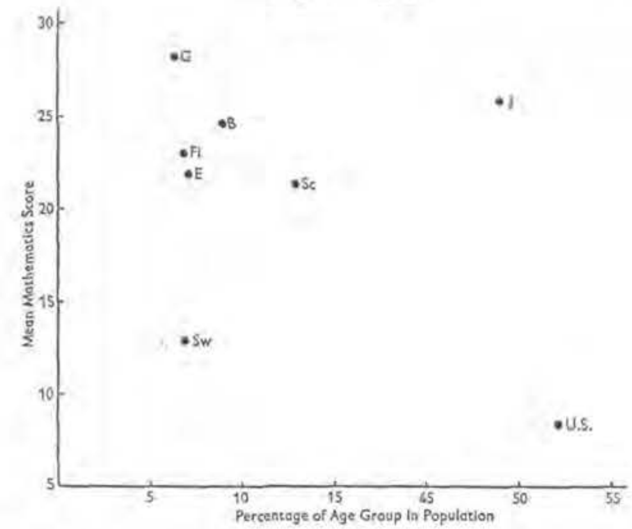


Fig. 5.2. *Relation of Mean Mathematics Score to Percentage of Age Group in Population by Country*

(Population 3 b)



the smaller would be the standard deviation, since those retained are likely to be more homogeneous in terms of mathematics achievement. There is some support for this, since the rank correlations between the percentage of an age group in school and standard deviation are .20 and .60 for Populations 3a and 3b respectively. The standard deviation is more likely to depend on how the groups retained are organized either within schools or between schools, and not just on the proportion retained. This must be a matter for further research.

Fixed International Standards Performance

Apart from examining the relationship between average scores and retentivity between countries, it is also interesting to employ another method of examining this problem — that of fixing a set of international standards to find what proportion of its pre-university students each country has been able to bring to each of these standards. Thus, we can examine not only what is achieved by the best students in each country, but also by the less able.

It has already been pointed out that there are major variations among the pre-university populations in the various countries in terms of some independent variables. With all these differences in mind, one might query whether it is justifiable to use combined distributions of scores from all countries as a base from which to derive percentiles for international comparisons. The reply would be that, whatever the national populations that contributed to produce them, the scores marked by the 95th and 85th percentiles of the combined distributions denote fixed points which can be used for at least some comparisons. For example, the 95th percentile for Population 3a is the score exceeded by only the best five percent of the combined pre-university populations for that level. If this five percent were composed of exactly five percent from each of the national pre-university populations, we should conclude that, in this respect at least, all the participating countries were equal. If the five percent international élite is not so composed, the question arises whether the differences are attributable, in part at least, to the varying percentages of the age group still at school.

Table 5.3 presents for each country the percentage of those students in Population 3a reaching six different *international* percentile levels.

Table 5.3. *Percentage of pre-university mathematics students reaching given standards.*

Country	Retentivity	International percentiles					
		25th	50th	75th	85th	90th	95th
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
U.S.A.	18	36	18	9	7	4.5	3.6
Sweden	16	81	53	26	13	8	3.1
Australia	14	67	37	10	5	3	1.1
Japan	8	82	63	43	29.4	21	10.0
Finland	7	81	48	18	6	3.4	1.2
Scotland	5.4	83	44	16	9	6	3.7
England	5	94	79	50	34	26	12.0
France	5	92	69	39	29.2	22	9.0
Netherlands	5	97	77	35	14	5	1.3
Fed. Rep. of Germany	4.7	90	63	26	11	7	2.0
Belgium	4	90	70	44	30	23	21.0
Range		61	61	41	29	23	19.9
Rank correlation with column 1		−.61	−.72	−.47	−.59	−.52	−.35

For example, 36 percent of the 3a Population in the U.S.A. reached the 25th percentile level, as compared with 97 percent of the 3a Population in the Netherlands. First decimal places have been added to some entries to increase the precision of the rank correlations. These rank order correlations between the percentage of an age group in Population 3a (i.e. Column 1 in Table 5.3) and the percentage of that population reaching each percentile level are shown in the last row of the table.

The negative correlations indicate that the smaller the proportion of the total age group taking the mathematics programme at the pre-university stage, the larger will be the proportions reaching given levels of performance. Thus, those who maintain that increasing the intake will lower the "standards" have a point, particularly in terms of the bottom half of those taken in. However, it is of interest that the effect at the upper end of the distribution is weaker. The between country ranges of percentages scoring above various international percentile points are very large, ranging from 61 percent at the 25th and 50th percentiles to 19.9 percent at the 95th percentile (see Table 5.3). Of those countries where only four or five percent of an age group are enrolled in the mathematics programme, Belgium and England are outstanding, particularly in the top international quartile. It is remarkable that 21 percent of Belgian students achieve scores above the 95th percentile (as, for example, compared with 12 percent in England) when it is remembered that Belgian students are studying an average of six more subjects than English students. The Netherlands, on the other hand, has a high proportion of students up to the 50th international percentile, but a rapid fall then occurs. The U.S.A. is consistently lower than Sweden (except at the 95th percentile), whereas Japan is consistently higher than Scotland (except at the 25th percentile).

If there were no relation between the degree of retention and the scores made by the students retained, we might expect that each country would have 5% of their 3a Population above the 95th percentile, 10% above the 90th percentile, etc. It will be seen from Table 5.3 that this is not the case. Countries with a higher rate of retention bring less than five percent to the 95th percentile. Although in general the less the intake the better the performance, there are some interesting differences among countries with similar enrolments. Scotland, England, France, Netherlands, Germany, and Belgium all have similar sizes of intake, but differ considerably in the proportions of the enrolment they bring to the international top three percentile levels.

Although the suggestion that "more means worse" has been seen to have some justification, in particular in the bottom half of the distribution, it is more meaningful to see whether the size of the "élite" group (as a proportion of the total age group) can be increased by increasing the size of the intake. If the numbers reaching particular percentile levels are calculated as percentages of the *whole age group*, some differences may become apparent. These percentages are presented in Table 5.4.

The rank order correlations between the percentage of an age group enrolled in the mathematics-science programme and the percentage of the whole age group reaching various percentile levels are given in the last row of Table 5.4.

Table 5.4. *Percentage of age group reaching given standards.*
(Population 3a)

Country	Retentivity	International percentiles					
		25th	50th	75th	85th	90th	95th
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
U.S.A.	18	6.5	3.2	1.6	1.3	.81	.65
Sweden	16	13.0	8.5	4.2	2.1	1.28	.50
Australia	14	9.4	5.2	1.4	.7	.42	.15
Japan	8	6.6	5.0	3.4	2.3	1.68	.80
Scotland	5.4	4.5	2.4	.8	.5	.32	.19
Finland	7	5.7	3.4	1.3	.4	.24	.08
England	5	4.7	3.9	2.5	1.7	1.30	.60
France	5	4.6	3.4	1.9	1.5	1.10	.45
Netherlands	5	4.8	3.8	1.7	.7	.25	.06
Fed. Rep. of Germany	4.7	4.2	3.0	1.2	.5	.32	.09
Belgium	4	3.6	2.8	1.8	1.2	.92	.84
Range		9.4	6.1	3.4	1.9	1.44	.78
Rank correlation with column 1		+.89	+.55	+.15	+.25	+.14	+.10

These positive correlations indicate that the higher the enrolment is as a percentage of the *total age group*, then the higher is the percentage of the whole age group reaching various international percentile levels. The greatest changes from Table 5.3 to Table 5.4 occur in Sweden, the U.S.A. and Japan, all three countries with a more retentive system at the secondary level. Thus, it is possible to increase the size of the élite group (as a percentage of the total age group) but only to a small extent.

Again, the between-country range varies from 9.4 percent at the 25th percentile to .78 percent at the 95th percentile. The percentage of the whole group reaching particular international percentile levels is obviously a function of size of enrolment to a large degree at lower levels though less so at the top levels. It is perhaps not without significance that students reaching the 99th percentile are drawn only from the U.S.A., Sweden, and England (.18, .16, and .05 respectively of their respective total age groups).

Performance of the élite group (in terms of the top ten and five percent international group), is weakly associated with size of enrolment. It is Japan, Sweden, England and Belgium which are performing well. Perhaps the significance of this finding becomes more apparent when phrased in another way: it would appear that countries with higher retentivity are capable of bringing their best pupils (in terms of the same percentage of a year group) to the same standards as less retentive (more selective) countries, i.e., higher retentivity does not necessarily mean lowering the standards of achievement (at least in mathematics) of the better students.

Similar information for Population 3b is given in Tables 5.5 and 5.6. The results agree closely with those obtained for Population 3a. There is a negative relationship (except at the 95th percentile) between the percentage still at school and the percentage of that population reaching various international percentile levels. The small size of the negative correlations for the 75th, 85th and 90th percentiles and the positive correlation at the 95th percentile indicate that at these levels, the degree of retentivity is irrelevant or, at the top level, favourable for high scores. Again, as with 3a, if the numbers reaching the various percentiles are calculated as proportions of the *total age group*, there are positive correlations.

Retentivity in the terminal mathematics-science programme is *negatively* related to the proportions of those still at school reaching

Table 5.5. *Percentage of pre-university non-mathematics students reaching given standards.*
(Population 3b)

Country	Retentivity	International percentiles					
		25th	50th	75th	85th	90th	95th
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
U.S.A.	52	30	12	3	2	1	1
Japan	49	81	60	38	28	21	12
Scotland	12.6	82	50	18	7	3	1
Belgium	9	93	63	27	15	8	2
England	7	84	53	20	10	5	2
Sweden	7	56	10	2	0	0	0
Finland	7	90	57	17	10	5	1
Fed. Rep. of Germany	6.5	99	81	37	20	8	1
Range		69	71	36	28	21	12
Rank correlation with column 1		−.99	−.28	−.02	−.09	−.06	+.38

various international percentile levels. Retentivity is *positively* related to the proportion of the *total* age group reaching various international levels. In general, the systems having smaller intakes of either 3a or 3b have achieved a fairly high performance of the weaker students in the programme. When an intake is increased in size, it is the performance of this lower group which tends to deteriorate. Nations can, however, certainly increase their total "mathematical yield" of an age group by having larger intakes (higher retentivity). In terms of the top international ten and five percents, retentivity is only weakly related to the proportions of the total age group reaching these levels, i.e., the performance of high ability students is unlikely to be affected by increasing the intake.

In Population 3a, Belgium, England and Japan have a consistently high performance of all students. Sweden and Japan demonstrate very well that increasing the size of the intake does not necessarily mean lowering standards. Sweden has an intake approximately three times larger than, for example, that of England, and yet approximately the same proportions of the total age group are still reaching 90th and 95th percentiles. Again, although systems with smaller intakes bring these students to higher mean scores, this is only to be

Table 5.6. *Percentage of age group reaching given standards.*
(Population 3b)

Country	Retentivity	International percentiles					
		25th	50th	75th	85th	90th	95th
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
U.S.A.	52	15.2	6.2	1.6	1.0	.52	.52
Japan	49	39.7	29.4	18.6	13.7	10.3	5.9
Scotland	12.6	10.3	6.3	2.3	.88	.38	.13
Belgium	9	8.4	5.7	2.4	1.3	.72	.18
England	7	5.9	3.7	1.4	.70	.35	.14
Sweden	7	3.9	.70	.14	0	0	0
Finland	7	6.3	3.9	1.2	.70	.32	.07
Fed. Rep. of Germany	6.5	6.4	5.3	2.4	1.3	.52	.06
Range		35.8	28.7	18.6	13.7	10.3	5.9
Rank correlation with column 1		.81	.95	.34	.40	.53	.81

expected when the selection processes and smaller numbers are considered. What is more important, however, is the proportion of the total age group reaching particular levels. Here the size of intake may have an important effect at the lower levels (see Table 5.4, Sweden at 25% level), and at the top levels it is possible for countries with large intakes (e.g. the United States and Sweden) to bring high proportions of an age group to the 90th and 95th international percentiles. At the top level Finland, Australia, the Netherlands and Germany are performing extremely poorly. Germany is particularly surprising, considering its high selectivity. From Table 5.3 it appears that the weaker half of the U.S.A. group is below the standards of other countries.

For Population 3b, Japan, Belgium and Germany perform well, whereas Sweden and the United States perform relatively poorly. It must be remembered that in Germany the 3b group have all studied mathematics up to the end of the penultimate preuniversity year (i.e. the *Unterprima*).

It is interesting to note those countries whose Populations 3a and 3b both perform well and those where there is considerable disparity. However, before arriving at any firm conclusions, it is necessary to bear several points in mind. First, there are differences between systems as to when students are allowed to discontinue the study of mathematics. Secondly, there are differences as to what discontinuing means; in some countries, it means absolutely no further mathematics and in other countries it means having mathematics for one or two periods a week instead of seven or eight periods a week. Thirdly, it must be borne in mind that the distinction between Populations 3a and 3b is somewhat circular, since, where it was difficult in some countries to distinguish between those pre-university students who were said to be specialising in mathematics and those who were not, a way of operationalising the distinction was to give the 3a tests to those groups of students for whom the tests were thought to be appropriate and then the 3b tests to the rest of the students.

Another approach to this same problem described in Husén *et al.* (1967) was to compare the performances of equal proportions of an age group; as a result, the same conclusion as above was reached, i.e., that the performance of the best three or four percent of students in a country is not affected by an increase in the intake (retentivity) into the pre-university year, but that the *average* score of all those in school in either the mathematics or non-mathematics programme will fall as the proportion of an age group retained increases.

Yield

As has already been pointed out in examining the "outcomes" of a system of education, it is often misleading to compare mean scores. It would be pointless to compare the mean score of the English students in the mathematics programme in the pre-university year with the United States students in the 12th grade mathematics programme. It is imperative to take into account the proportion of an age group still studying mathematics, i.e., *"how many of these students are brought how far?"* For example, in England only five percent of an age group is studying mathematics in the pre-university year, whereas in the United States eighteen percent of an age group is studying mathematics at that point.

There are difficulties connected with the calculation of a "yield" or "output" measure. A simple statement of the overall problem is "How are achievement scores and number of students having a given score to be combined into a single measure of output?" Two very

simple approaches are used here. The first consists of plotting the cumulative percentile frequencies (or percentile frequencies could be used) against the percentage of an age group in a particular target population and regarding the area under the curve as the "yield". The second consists of multiplying the proportion of an age group in a target population by the mean score of the population and regarding the resultant value as an index of "yield".

The difficulties with these approaches are best exposed by considering the assumptions behind them:

(a) Each correct response to an item is regarded as being of equal value. Thus, two students having the same scores are regarded as representing the same output even though one student may have correct responses on items which are considered to be either more difficult or of more value to society than another student.

(b) Each point on the achievement scale has the same absolute value as every other point. Thus, the increment from 23 to 24 represents the same increase in "output" or "yield" as an increase from 40 to 41. It is, of course, possible that, in some case, 20 points may be twice as valuable as 10, and, in another case, 40 may be less than twice as valuable as 20.

(c) One student with a score of 20 is considered equal in terms of yield to two students with scores of 10 each.

(d) The value of the *n*th unit of achievement is assumed to be the same in all countries, although countries may differ in their economic structure. This, however, introduces the concept of "required (by the society) yield" and its fit to "acquired yield".

Despite the problems involved in calculating "yield", the simple approaches mentioned above will be presented since the concept of "yield" or "output" is important. As has already been mentioned in Chapter 1, what is reported here are the yields of specific target populations. To obtain a measure of the "total yield" of a school system, the achievement of all those dropping out of school has to be measured as they drop out and in some way brought into a single measure. A longitudinal approach could also be adopted.

The yield of students in Population 3a will be examined first, followed by that of the total pre-university year (Populations 3a and 3b combined) and finally the yield of 13-year-olds will be compared with the pre-university yields in each country.

Population 3a

Figure 5.3 represents the yield diagrammatically by plotting the cumulative percentile frequencies for each country against the proportion of an age group still retained in the terminal mathematics-science programme. These distributions have been smoothed graphically. From Figure 5.3 it can be observed that it is Sweden, the United States, Australia and Japan which have the highest yields, despite the fact that in the first three countries the average scores were relatively low. Obviously, yield is, to a certain extent, a function of retentivity, but *only to a certain extent*. The United States' yield is obviously smaller than those of Sweden and Australia, although the United States' retentivity is higher.

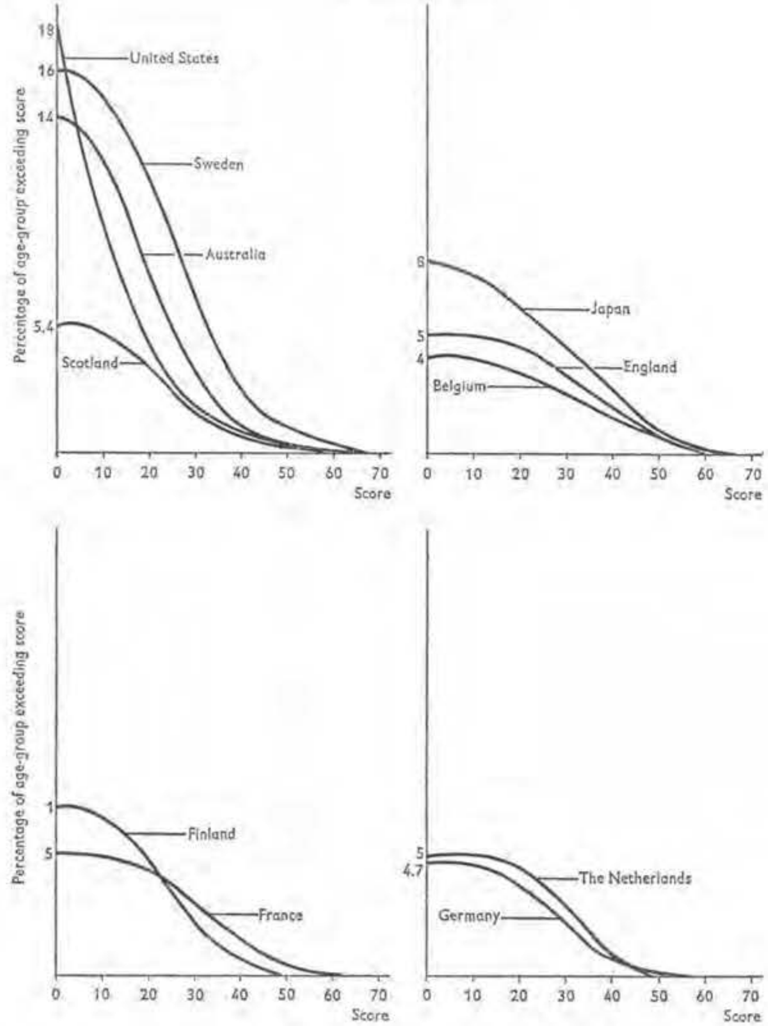
It is interesting to note that in some countries there is a consistently higher performance over the whole range of students than in others (e.g. Japan as compared with Finland). The United States' students at the lower and of the distribution perform less well than the Swedish students. French and English students perform relatively well at the top end of the distribution.

Population 3a and 3b

Although it is only Population 3a which can be regarded as the mathematical "fruits" or "end-products" of a system of education, it is also of interest to examine the yield of Populations 3a and 3b together, since this comprises the total proportion of an age group still in *full*-time schooling. What the yield would be of a total age group is a matter of pure speculation, since in this study no effort was made to measure the mathematics performance of those students in part-time education (and here the proportions of an age group in part-time schooling, whether compulsory or voluntary, differ considerably from country to country) or those young people of the age group not receiving any form of schooling. For example, in England there is a small proportion of an age group which studies pre-university mathematics at Colleges of Further Education or Technical Colleges, but such students were excluded from the target population. In the Federal Republic of Germany a considerable proportion of young people attend *Berufschulen* and continue the study of mathematics there. Again, these students were excluded from the target population, since they were not in *full*-time schooling. Thus, the

Fig. 5.3. *Cumulative Percentile Frequencies (Smoothed)*

(Population 3 a)



"yield" examined here is simply that of all pre-university students in the target populations.

Table 5.7 presents for each country the corrected mean score for Populations 3a, 3b and 1b along with the proportion of the age group still retained in school for each of these populations.

Table 5.7. *Total mathematics score and proportion of age group in school.*
(Populations 1b, 3a and 3b)

Country	Population 1b		Population 3a		Population 3b	
	Mean	Proportion	Mean	Proportion	Mean	Proportion
Belgium	30.4	100	34.6	4	24.2	9
England	23.8	100	35.2	5	21.4	7
Fed. Rep. of Germany	25.4	100	28.8	4.7	27.7	6.5
Finland*	16.1	100	25.3	7	22.5	7
Japan	31.2	100	31.4	8	25.3	49
Scotland	22.3	100	25.5	5.4	20.7	12.6
Sweden	15.3	100	27.3	16	12.6	7
U.S.A.	17.8	100	13.8	18	8.3	52

* Although the mean for Finland is given as 16.1 the *scaled* means (and yields) were calculated on uncorrected Finnish data where the mean was 26.4.

However, since Test 5 was common to both Populations 3a and 3b it was possible to estimate[3] what the 3a students would have scored on the 3b tests had they performed in the same way as they did on Test 5. Furthermore, since Test 3 was common to Populations 3b and 1b, it was possible to estimate what 1b students would have

[3] A regression procedure was used for each country to predict a test 3 score (t_3) from the total level 1b score (T_{1b}) and then predicting from that t_3 to an estimated T_{3b} on the 3b scale. The two regression equations were:

$$t_3 = a_1 + b_1 T_{1b}$$

and

$$\hat{T}_{3b} = a_3 + b_3 t_3$$

which combine to give

$$\hat{T}_{3b} = a_3 + a_1 b_3 + b_1 b_3 T_{1b}$$

where

$$b_1 = \frac{N \Sigma T_{1b} t_3 - (\Sigma t_3)(\Sigma T_{1b})}{N \Sigma T_{1b}^2 - (\Sigma T_{1b})^2}$$

$$a_1 = \bar{t}_3 - b_1 \bar{T}_{1b}$$

$$b_3 = \frac{N \Sigma t_3 T_{3b} - (\Sigma t_3)(\Sigma T_{3b})}{N \Sigma t_3^2 - (\Sigma t_3)^2}$$

$$a_3 = \bar{T}_{3b} - b_3 \bar{t}_3$$

The same procedure was used for reducing the 3a to 3b score.

Table 5.8. *Correlations of tests 5 and 3 with total mathematics score.*

Country	Population 3a	Population 3b		Population 1b
	Test 5	Test 5	Test 3	Test 3
U.S.A.	.91	.86	.90	.79
Japan	.90	.94	.91	.90
Sweden	.86	.73	.80	.78
Scotland	.82	.87	.88	.87
Finland*	.84	.85	.84	.78
Belgium	.86	.86	.85	.85
England	.88	.87	.88	.90
Fed. Rep. of Germany	.78	.80	.79	.82
All Countries	.89	.92	.91	.86

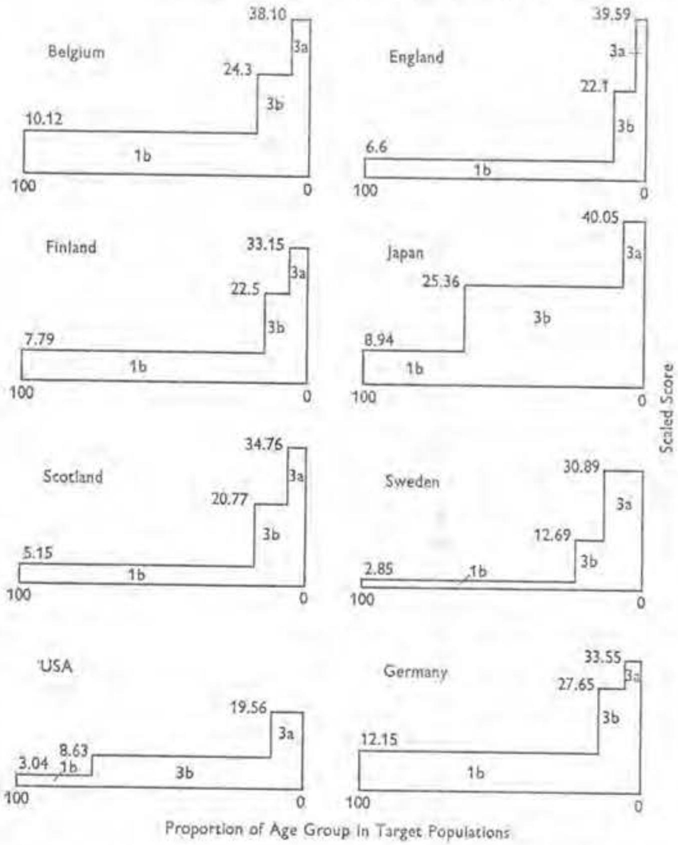
* These correlations were calculated on the uncorrected Finnish data.

scored on the 3b tests had they performed in the same way as they did on Test 3.[3] However, it must be remembered that the content of 3a and 3b tests differed considerably from Test 5 and also the content of the 3b and 1b tests from Test 3, as can be seen in the Appendix to Volume II of Husén *et al.* (1967); this accounts for the differences between Table 5.7 and Figure 5.4, where scaled means are given.

Table 5.8 presents the product moment correlation coefficients between the Total Mathematics Scores (corrected) and Test 5 and Test 3 scores (as the case may be) for each population in each country. The Total Mathematics Score included the Test 5 (or 3) scores and hence the correlations are higher than if it were Test 5 (or 3) scores correlated with the Total Mathematics Score minus Test 5 (or 3).

Figure 5.4 presents the diagrams of scaled means for Populations 3a, 3b and 1b against the proportion of an age group still in school for each of these populations. Each diagram is made up of three parts as follows. The base of each diagram consists of the 1b population (where 100 percent of an age group is estimated to be in full-time schooling); the proportion of an age group is shown on the horizontal axis and the scaled mean score on the vertical axis. A similar procedure is used for the 3b population and for the 3a population shown at the right side of each diagram.

Fig. 5.4. *Combined Yield (3a + 3b on 1b)*
(Regression Scaling of 3a + 1b on to 3b)



In Figure 5.4 the effect of retentivity on yield can be seen. Japan has a particularly large yield. It should, however, be remembered that the procedure used here does not take into account those students who have left school between Populations 1 and 3.

It is possible to calculate a yield coefficient for each population by multiplying the scaled mean (or ordinary mean) by the percentage of an age group in school. The percentage of an age group in school for the 1b population is estimated to be 100% in each country. The combined yield of the pre-university year is the sum of the yield coefficients for Populations 3a and 3b. These yield coefficients are given in Table 5.9 (the scaled means are given in Table A.4 in the Appendix).

Table 5.9. *Yield coefficients.*

Country	On scaled means				On ordinary means			
	1b	3a	3b	3a + 3b	1b	3a	3b	3a + 3b
U.S.A.	304	352	449	801	178	2484	4316	6800
Japan	894	320	1243	1563	312	2512	12397	14909
Sweden	285	494	89	583	153	4368	882	5250
Scotland	515	188	263	451	223	1377	2608	3985
Finland	779*	232	157	389	161	1771	1575	3346
Belgium	1012	152	219	371	304	1384	2178	3562
England	660	198	155	353	238	1760	1498	3258
Fed. Rep. of Germany	1215	158	180	338	254	1354	1800	3154

* The scaled means were calculated on the uncorrected Finnish data. It has not been possible to rerun the regression scaling analyses since the mistake in the Finnish data was discovered.

The rank correlations between the scaled mean yield coefficients and the ordinary mean yield coefficients are .79, .91, 1.0 and .98 for 1b, 3a, 3b and 3a plus 3b respectively. The correlations indicate that there is a high degree of relationship between the two types of means used to calculate the yield coefficients. In terms of the pre-university yield (3a + 3b) it is worthy of note that although the United States has three times as many pupils as Sweden enrolled in the pre-university year, its yield is only 25 percent greater. Again, Japan has just over twice as many pupils as Sweden enrolled in the pre-university year, but has a yield nearly three times as great.

It is of particular interest, when considering yield, to compare the yield of the 1b population with the pre-university yields. Since the 13-year-old grade group was the last point in all the school systems where 100 percent of an age group was still in school, it can be considered as a comparable point near to the end of compulsory schooling, and the yields as fairly representative of the outcomes of the compulsory schooling in each country. At the same time, it must be realised that the actual age of ending compulsory schooling differs from system to system and that some countries will obviously increase this yield before the end of compulsory schooling.

It seems likely, for example, that, in those countries where compulsory schooling does not end until the age of sixteen, certain topics which are considered to be difficult may be postponed until the age of fifteen, while in those countries where compulsory education finishes at fourteen years of age, these topics may be introduced at thirteen years of age. It might have been better to use Population 1a instead of 1b for these yields, since this is a strictly chronological group, but as pointed out earlier in this chapter, this would have provided results for seven countries only, since Germany did not test Population 1a. Therefore, despite the limitations involved, it was decided to use Population 1b.

The rank correlation between the 1b yields and 3a+3b yields is –0.56. Germany and Belgium are particularly worthy of note here, since from the 1b yield to the 3a+3b yields they move from first and second places to last and 6th respectively. Only the United States, Japan and Sweden have relatively higher yields at the pre-university level than at the 1b level and this is obviously, to a certain extent, a function of the size of retentivity. It would seem that the less retentive systems lose a great deal of potential mathematical knowledge in their countries, and, at the same time must also lose a certain amount of talent. The rank correlation between yields (scaled) for Populations 3a and 3b (separately) and the measures of social bias[3] given in Chapter 3 of Volume II of Husén *et al.* (1967) are +.56 and +.56. (The measures of social bias are repeated in Table A.5 in the Appendix to this book).

Thus it can be seen that the pre-university yield is negatively related to the 1b yield, but is positively related to social bias which is in turn related to the age at which selection takes place (see Husén *et al.*, 1967). At the same time, we know that yield is, to a certain extent, a function of retentivity and retentivity is related to the percentage of pupils in Population 1a in comprehensive schools (see p. 67). It would seem that in countries with higher yields at the pre-university level, there is a philosophy of equality of opportunity in that selection is delayed or abolished, comprehensive schools are more common and more pupils from lower social status families continue through to the end of secondary schooling.

These organizational features, however, are not alone responsible for high yields, as seen by the difference between the United States', Japanese and Swedish yields. The curriculum, teaching and other

[3] Social bias is an index of the degree of difference of the socio-economic composition of one group to another, in this case Population 1a to 3a and 1a to 3b. It can be reasonably assumed that Populations 1a and 1b have nearly identical socio-economic distributions.

family background characteristics are the most likely factors to account for other differences (see Chapter 6 of Volume II in Husén *et al.*, 1967).

Although some factors associated with yield have been examined, no mention of the relationship between this yield "acquired" by the systems and the yield "required" by a society has been made, since it is not known. Research similar to that carried out by Dahllöf (1963) would have to be undertaken where different branches of society receiving students from school could estimate the amount of knowledge they require from these students in a particular subject, and where, at the same time, approximations could be made of the proportion of any one age group entering work in that branch of society. In this way, it would be possible to estimate the "required" yield. Yield, as discussed in this chapter, has been based on Total Mathematics Score; it would, of course, be possible to discuss yield in terms of particular topics in mathematics and clusters of topics. By comparing "required" with "acquired" yield, it would be possible to examine how well the schools prepare their students to meet the needs of the society. This is *not* to imply that a school system should be based on a purely utilitarian philosophy; it should, of course, have much wider aims. Nevertheless, one of its basic tasks should be to meet the needs of the society. At present, however, the only system, to the author's knowledge, where this problem of "required" yield has begun to be examined empirically is Sweden. In other countries, there is only intuitive knowledge of what society requires.

Although it is possible to obtain ratings of the amount and type of mathematical knowledge required by various sectors of the society (including the university) receiving students straight from school, the problem becomes difficult when prediction in terms of manpower requirements with certain mathematical competences is attempted —the concept of "fit". This is so because, in the economist's language, "demand" is never a fixed amount but rather a schedule. Furthermore, the principle of substitution operates so that to some extent x "poorer" mathematicians can be substituted for y "better" mathematicians. Thus, the question becomes that of how many mathematicians are desired at each alternative price per unit. Added to this is the problem of predicting future demands. What is self-evident is that in the application of the concept of "fit" an interdisciplinary attack is required.

Summary

It has been shown, by a discussion of the relationship between retentivity and type of school system, how the traditional European system, involving selection into an academic secondary school, has a lower rate of retentivity than the United States' system with its self-contained establishments which have been continually expanded to satisfy the educational needs of the community. However, in many of the European countries at present, policies concerning school structure are being revised, and in Sweden, for example, the percentage of seventeen-year-olds proceeding to *gymnasiet* has risen from ten to 28 percent from 1950 to 1964 (Dahllöf *et al.*, 1966).

The percentage of an age group in both the pre-university mathematics and non-mathematics programmes in the twelve countries in the study was related to an index of comprehensive schooling in the various countries and the correlations obtained were high. Three examinations of the pre-university scores were made in connection with retentivity: an examination of average performance of the various populations, an examination of the performance of the various populations at fixed international standards and an examination of the "yield" (how many are brought how far) of the various populations. Variations among the pre-university populations in such characteristics as age, social class composition and number of subjects studied are pointed out.

The examination of average performance shows that countries which retain larger percentages of an age group to the pre-university stage produce on *average* lower standards of achievement than do countries retaining smaller percentages. However, the range of scores was not related to retentivity, although this would have been expected.

In the examination of performance at various fixed international standards it becomes clear that although the *average* score may drop when a higher proportion of an age group goes through to the pre-university year, the performance of the best students (in terms of the proportion of a year group reaching various international percentile levels) does not necessarily deteriorate. In other words, an increase in intake into the pre-university year does not necessarily cause a drop in the levels of achievement of the best students. This finding is of particular importance in the light of the fears of many teachers who argue that if more and hence poorer students are allowed through, the standards of performance will deteriorate and the learning of the better students will suffer.

Since this is the case, it is interesting to proceed to an examination of the "yields" (how many are brought how far) in mathematics of the pre-university populations in the eight countries which had scores for both Populations 3a and 3b. "Yield" takes into account the differing proportions of an age group in these populations in the different countries, whereas a comparison of *average* performances of pre-university year students in different countries does not. A diagrammatic presentation of "yield" for Population 3a is given, and this is also given in terms of "yield coefficients" (calculated on scaled mean scores as well as ordinary mean scores) for both Populations 3a and 3b. In general, systems with higher retentivity have greater yields, but yield is, to a certain extent, a function of retentivity. Curriculum, student motivation and other factors also would seem to play some part in accounting for other differences in performance. It would seem that further research is needed to explore these issues. The relationship between Population 1b yields and the pre-university yields was negative and is mainly, but not entirely, due to the varying retentivity through to the pre-university year. It would seem that in some countries, particularly Germany and Belgium, a great deal of talent drops out of regular full-time schooling. This is, in turn, related to the selection process in some countries and results in bias in the social status composition of the students in the pre-university years in favour of the higher social status groups. The data obtained in this study reveal clearly the possibility of having both a high overall yield and an undiminished élite yield.

Although the concept of "yield" or "output" introduced is somewhat crude, it is an important one and it is to be hoped that its conceptualisation and operationalisation will be pursued, and that it can be so refined in the future to produce detailed measures of "acquired" yield in many subject areas. Measurement of "required" yield has already been begun in some areas. When progress is made in the measurement of the types of yield—that produced by the school system and that required by society—it will be possible to compare them and although the concept of "required yield" has its difficulties, the whole notion of "fit" may provide the schools (and educational policy makers) with more insight into the ways and means of catering for the needs of society.

CHAPTER 6

Differentiation

Three different aspects of differentiation will be examined in this chapter in the light of the data available from the IEA study. The focus will be on the range of performance in systems employing varying modes of differentiation. In terms of inter-school grouping, some countries have a selective system whereby the more able students at a particular age are separated from the main body of students and put into selective-academic schools; other countries have a comprehensive system in which all students are kept in one school type until the end of compulsory schooling or until the end of secondary schooling. This is what is meant by differentiating or not differentiating into different school types, and is sometimes referred to as *organizational differentiation* (cf. Husén, 1962 *a*, and Yates, 1966). An examination will be made of the range of mathematics scores of students in the grade where most 13-year-olds are to be found (Population 1b) in comprehensive and in selective systems of education. In addition, there is intra-school grouping, which concerns the grouping of students within schools—sometimes referred to as *educational differentiation*. Some countries have a system of grouping students by grades with promotion taking place on the achievement of a certain standard; other countries promote by age groups. In the first case there is often a sizeable proportion of an age group not in the grade group in which most of the age group would be found if promotion were by age. The amount of retardation varies from country to country. In the second case nearly a hundred percent of an age group are in one grade. An examination of the range of mathematics scores of 13-year-old students and of the corresponding grade group will be made from both the grade promotion systems and the age promotion systems. A further form of educational differentiation is that of streaming or ability grouping, whereby students are split into different groups within a grade on the basis of measured or judged ability and/or achievement. The extent to which this is carried out varies a

great deal from country to country. This is the third aspect of differentiation to be considered in this chapter and will involve an examination of the range of mathematics scores of students in Population 1b from countries where ability grouping is practised to a great extent and from those countries where it is practised either to a small extent or not at all.

A great deal of research has been carried out on various aspects of differentiation and particularly into ability grouping. In recent years various summaries of the research carried out have been made (cf. Ekström, 1961, Goldberg *et al.*, 1966 and Yates, 1966) and these include all of the research studies which are relevant to the three aspects of differentiation described above. Most of the research so far carried out can only bear very peripherally on the problems under discussion here, and the directly relevant (in that the standard deviation scores have been used as a criterion) studies are very few indeed. Svensson (1962) carried out a five-year follow-up study where he compared the performance of students under a comprehensive system of education and students under a selective system of education in the City of Stockholm from 1955–59. His findings were that by the age of fifteen, "good" students performed at about the same level whether in the selective-academic school *(realskolan)* or in the comprehensive school *(grundskolan)*, whereas "poor" students performed better in the comprehensive school than in the remainder school *(folkskolan)*. Although Svensson did not specifically compare standard deviations, the implication is that the standard deviation is smaller in the comprehensive than in the selective system (when the performances of students in different schools are combined). In an article by Husén and Svensson (1959) and from certain findings in Chapter 3 of Volume II of Husén *et al.* (1967), the same implications are apparent. There is other research which, although it does not compare selective and comprehensive systems, shows how streaming influences the standard deviation within a school type. Douglas (1964), has followed the complete population born in the first week of March, 1946, right through their school careers; this follow-up is still continuing. It became apparent that when children were tested or assessed on the basis of ability for placement into higher or lower academic groups (whether this was within schools or between schools), those who entered the higher academic groups were more frequently from the higher social status groups and these students continued to improve; on the other hand, those who went into the poorer groups were often from the lower social status groups and their performance over a period of time deteriorated relative to the higher social group. Even when children at age 8 had the same score, it was the middle social status group children who tended to be put into the higher group, while the lower social group children were placed in the lower group. Certain analyses which appeared in the Robbins Report (1963) are a follow-up of the information in Douglas' book and indicated that the trend which he had already detected up to the age of 11 continued for students going on to 15 and 18. Pidgeon (1959) has shown in a national survey of attainment in mechanical arithmetic the percentage of modern school and all age school children scoring above the grammar school mean was 22% at age 14, i.e., that despite selection at age 11, there was still a very big overlap of scores between the secondary academic school students and the remainder of the students. This may well reflect the limited range of the content of the tests, but on the other hand, it may be indicative of different rates of development in the whole range of children, with the result that the modern school does not necessarily possess the weaker children at all levels. Since grouping between schools by ability/achievement is based on the same principle as streaming, it seems reasonable to infer that selective systems which also practise streaming will have the largest standard deviations of all systems. Pidgeon (1962) examined the concepts of streaming versus non-streaming and grade promotion versus age promotion in terms of the standard deviation of 13-year-olds in twelve countries. It is clear from Pidgeon's data that selective systems do not necessarily have larger standard deviations than comprehensive systems, but it must be remembered that this study was carried out on 13-year-old samples of students, the representativeness of which was unknown.

A number of other studies have questioned certain aspects of inter-school grouping based on differences in ability and attainments. Yates and Pidgeon (1957), Emmett (1945), Daniels (1959) and others in Britain, as well as Hitpass (1960) and Undeutsch (1960) in the Federal Republic of Germany have shown that even the best available methods of allocation involve errors of placement with regard to at least ten percent of the children concerned. Pedley (1963) and Dancy (1963) in Britain have shown that students who would not

normally have entered grammar schools have proved capable of grammar school type success from comprehensive or independent schools. The fact that this is remarked upon indicates that there is thought to be a gap and the implication is that if all were educated together the gap (and hence the standard deviation) would be smaller. This reinforces the view that educational systems practising inter-school grouping are expected to have larger standard deviations than countries not practising it.

As far as age promotion versus grade promotion is concerned, there is no known research. Belgium, in its official statistics (1960–61), has published a table revealing the progressive increase in the incidence of backwardness as children move through successive grades.

Grade	1st	2nd	3rd	4th	5th
% of students of normal age or above	84	77	74	71	69
Index of school backwardness	24	35	41	45	47

An index of the amount of grade repeating and grade advancement in any country will be the size of the standard deviation of age of students in Population 1 b. These are given in Table 6.1. As can be seen, England, Japan and Sweden have the smallest standard deviations, while the Netherlands and Belgium have the largest. In England, a system of grades (known as "standards") used to operate, but has largely been abandoned in favour of what is sometimes known as horizontal grouping, which involves promotion by chronological age. In Sweden, chronological age is the basic criterion of grouping, although a certain amount of grouping based on subject-ability also took place from Grade 7 onwards. In most of the other European countries, however, and in the United States, some form of grading is practised. In Israel, on the other hand, the general practice of allowing (or requiring) a slow student to repeat a grade was recently discontinued and teachers are now asked to restrict non-promotion to two percent of their students. In the United States, more radical departures from the normal type of grading are being tried, and these are lucidly described in Goodlad and Anderson (1963) and in Rasmussen and Prete (1962).

A great deal of research exists on the form of educational differentiation involving streaming or ability grouping. Firstly, it must be realised that differentiating by ability either between or within

Table 6.1. *Means, standard deviations and N's of total mathematics score and standard deviations of age in months.*
(Population 1b)

Country	Total mathematics score M	S.D.	N	Age S.D.
Australia	18.88	12.28	3079	7.7
Belgium	30.43	13.75	2644	8.8
England	23.76	18.53	3148	4.2
Fed. Rep. of Germany	25.45	11.70	4476	6.6
Finland	16.13	11.61	1325	6.66
France	20.96	13.23	344	7.8
Israel	32.29	14.67	3232	5.6
Japan	31.16	16.90	2050	3.4
Netherlands	21.43	12.12	1444	11.6
Scotland	22.31	15.69	5718	5.4
Sweden	15.26	10.83	2808	4.9
U.S.A.	17.85	13.21	6544	6.8

schools is based on the same principle, and therefore much of the research already mentioned concerning inter-school grouping is relevant also to the problem of intra-school grouping. Yates (1966) has abstracted about 40 researches dealing with aspects of homogeneous grouping, which had been undertaken between 1932 and 1965. It is interesting to note that whereas the research into inter-school grouping, although sparse, has been fairly conclusive, the research into intra-school grouping, although plentiful, has been conflicting. Passow (1962) has described some of the discrepancies in the research so far undertaken which may well account for these apparent contradictions. The general findings of comparisons of homogeneous and heterogeneous groups or of streamed and unstreamed groups have mainly concentrated on differences in mean scores between the groups. However, from the work of Blandford (1958), Rudd (1958), Khan (1954), Gatfield (1958) and Daniels (1961) in Britain, one result of the comparisons, which is relevant to the present discussion, was "The dispersion of the various test results was greater in the streamed than in the unstreamed schools." (Yates, 1966, page 63.) This is to be expected, since in a heterogeneous group the teacher is likely to teach to a mean level with the result that the variance of scores will become less, whereas if a group is split into "n" homogeneous groups, then the variance of the group as a whole will increase. Pidgeon (1962) has suggested that much of this is bound up with teacher expectation and student role fulfillment. If streaming takes place and a group is split on the basis of ability and achievement into three sub-groups—an A class, a B class and a C class—the teacher having the A group will expect that group to do well; the students themselves will expect to do well; they will in fact do so. The contrary will be true for the C class. Thus, the variance will increase. Furthermore, it is clear that the earlier this process of streaming begins in a school, the more the variance will increase as students progress through the school (cf. Douglas, 1964, and the Robbins Report, Appendix I, pp. 46–52). This phenomenon also influences teachers' philosophy concerning the "capacity theory of intellect"—the assumption that every child has a limited and measurable ability—since streaming tends to make this a self-fulfilling prophecy.

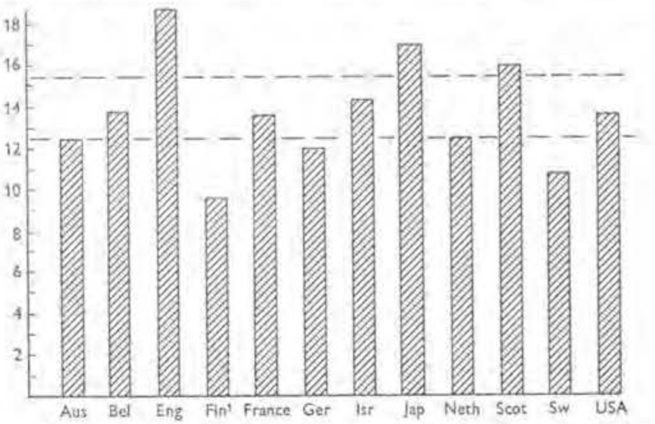
Within a year group setting (the grouping of students for specific subjects or activities only according to their ability or achievement) will have similar effects to streaming in increasing the spread of scores of the age group on any achievement criterion.

Let us now examine these three aspects of differentiation in terms of the data available from the IEA study for Population 1b (and 1a where appropriate). Population 1b has been selected for detailed examination, since it is a grade population within the limits of compulsory school attendance in all countries.

Inter-School Differentiation

The examination of the standard deviations of scores of a grade group from systems of education practising "organizational" differentiation of different extents will ultimately involve taking into account the amount of retardation (grade-repeating, etc.) in each of the systems (although this is already overlapping with the examination of the second aspect of differentiation). Table 6.1, therefore, presents not only the mean, standard deviation and number of students for Population 1b in each country, but also the standard deviation of the age of this population, since this can serve as an index of retardation in the system. (A full presentation of the means, standard

Fig. 6.1. *Standard Deviations of Mathematics Scores for 1 b Populations*



[1] The unweighted standard deviation for Finland is 11.61—see page 6 and also Table 6.1.

deviations and numbers of students of Total Mathematics Score, Lower Mental Process and Higher Mental Process by sub-sample appears in Table A.6 of the Appendix).

Figure 6.1 presents the standard deviations of mathematics scores diagrammatically.

From the presentation of the school structures in Chapter 4, it can be seen that Australia, Japan, Sweden and the United States can, on the whole, be placed in the theoretically non inter-school differentiation category, whereas the other countries have various degrees of inter-school differentiation. On examining Figure 6.1, it is evident that factors other than just inter-school differentiation are associated with the different sizes of the standard deviations. It is perhaps worth noting that the average standard deviation for differentiating countries is 13.65 and for non differentiating 13.33 ($p<.01$). However, it is obviously necessary to examine this in more detail. It is possible to split the countries into three groups: (1) those where the standard deviation is greater than 15.5 (2) those between 12.5 and 15.5 and (3) those under 12.5. In the first group are England, which has a selective academic system, Japan, which has a non-differentiated system, and Scotland, with a sizeable number of comprehensive schools. The standard deviation for England is significantly larger than that for Japan and that for Japan larger than that for Scotland. It was expected that England would have the largest standard deviation, since it practises not only inter- but also intra-school differentiation (streaming). Japan is a paradox—a system of mass education exists (57 percent of an age group still in school in the pre-university year), but although a junior high school and senior high school structure exists, it would appear that within these groups there is a hierarchy of schools (King, 1965) and there is severe competition among students to get into the best schools. This in itself already indicates a very severe form of inter-school differentiation with the best schools taking the best students and the poor schools having the poor students; this is likely to create a wider spread of scores than inter-school differentiation alone as practised in England without streaming. The gaps between the blocks of schools in England will be considerable, but the total range of between school differences is likely to be less than in Japan. At the same time, there is very little spread within Japanese schools, since it would appear (from discussions with Japanese educators) that motivation for learning is imposed by the teachers and that there is little in the way of structured content with motivation inherent in the learning situation. Thus, it seems possible that it is the hierarchy of schools which is associated with a wide spread in this case. (It would be possible to check this by a between-schools analysis). Scotland, although having more than half of its schools designated as comprehensive, practises a high degree of streaming within schools. At the same time, there are many small schools at the primary level which would tend to produce a fairly wide spread of scores.

In the second group are Israel, Belgium, the United States and France. Israel has a student population of wide ethnic background, often coming from countries with widely differing standards of education; in other words, the population was very heterogeneous and one of Israel's policies has been to try to homogenise the school population more and reduce the spread of scores (cf. p. 31).

On the other hand, all students who had immigrated to Israel after 1957 were excluded from the testing so that it could be argued that a smaller standard deviation might have been expected. As part of the homogenising policy an eight year elementary school now exists with transfer to secondary school taking place at the age of fourteen. Belgium and France, on the other hand, have the traditional European type of inter-school differentiation, without streaming, but with grade repeating, both to a considerable degree. The United

States, although not possessing *de jure* inter-school differentiation at the junior high school level, has *de facto:* a certain amount is found in the form of segregated schools in some areas; furthermore, ability grouping and enrichment programs are fairly commonplace (from a representative sample of Junior High Schools 66% of school principals said that in their schools ability grouping was practised universally or generally—Husén *et al.*, 1967).

Again, in the United States, students attend a school near where they live; since families of similar socio-economic status tend to live together, this has a homogenising effect on the schools in particular areas, e.g. suburbs, slums, etc.

In the third group are the Netherlands, Australia, the Federal Republic of Germany, Sweden and Finland. The Netherlands practises inter-school differentiation but differs from the other European selective systems represented in this study in that it is a system with a middle school. Definitive transfer to the academic-selective or pre-university school is not made at the end of the primary school course, but is deferred until the age of fourteen. The intervening period is spent in a common secondary school. However, grade repetition is practised in the Netherlands to a greater extent than in any other system in this study (see standard deviations of age in Table 6.1). Australia, although having a more or less comprehensive system of education, practises grade repeating and also ability grouping (see Table 6.3). Germany (and it must be reemphasized that the data representing Germany come from only two of its Länder—Hessen and Schleswig Holstein) has inter-school differentiation, no within school differentiation and a certain amount of grade repeating. Sweden has officially neither inter- or intra-school differentiation at this level (7 *årskurs*), although some within school differentiation takes place in Grade 8 and following grades. Finland practises inter-school differentiation, a certain amount of grade repeating (rank 5 in Table 6.1) and intra-school differentiation.

The above brief descriptions have served two purposes. First, they have attempted to supply qualitative descriptions of not only the inter-school differentiation which takes place, but also of the intra-school differentiation in terms of both grade repeating and ability grouping or streaming, which will be examined empririically later in this chapter.

Unfortunately, it has not been possible to consider the inter-school differentiation empirically because of lack of objective measure of the extent to which inter-school differentiation takes place in each system. It can, indeed, be seen from the above description how difficult it would be to establish an index for the type of *de facto* inter-school differentiation which exists, for example, in Japan. One possible measure on which data exist would be the retentivity index used in Chapter 5 whereby high retentivity could be regarded as analogous to little inter-school differentiation. Unfortunately, this would place Japan as having less inter-school differentiation than Sweden, which is obviously untrue. [If the total percentage of an age group retained to the pre-university year is used as an index of inter-school differentiation (low retentivity equivalent to high inter-school differentiation) the rank correlations between this and the standard deviation of mathematics scores for Populations 1a and 1b are .20 and −.57 respectively–Israel omitted from 1b–which does not accord with common sense.] However, this measure has too many limitations to be used in further analysis. It is clear that in future international educational research more thought must be devoted to developing a measure for this elusive variable. The measures obtained in this study of grade repeating and ability grouping are less limited.

Intra-School Differentiation—Grade Repeating

Table 6.1 presented the standard deviations of Population 1b students' Total Mathematics Scores and also the standard deviations of age in each country which serves as an index of grade repeating. The correlation between them is −.53, indicating that the more grade repeating is practised, the narrower is the spread of scores. This supports the theory that when a grade system of promotion is a feature of the system of education, then teachers will tend to teach to what they judge to be a mean level, which tends to reduce the spread of scores. In age promotion systems, the spread will be wider, since there will be a tendency either to allow students to progress at their own rates through the various subject contents to be learned, or to introduce ability grouping.

It is also of interest to examine the corresponding data for Population 1a. Table 6.2 presents the mean, standard deviation and number of students of Total Mathematics Score for each country as well

Table 6.2. *Means, standard deviations and N's of total mathematics score and standard deviations of age in months.*

(Population 1a)

Country	Total mathematics score M	S.D.	N	Age S.D.	Age S.D. (1b)
Australia	20.18	14.01	2916	3.5	7.7
Belgium	27.74	15.02	1686	3.3	8.8
England	19.31	16.97	3012	3.3	4.2
Finland[a]	15.39	10.76	1156	3.3	6.7
France	18.32	12.37	2410	3.5	7.8
Japan	31.16	16.90	2049	3.4	3.4
Netherlands	23.86	15.91	428	3.1	11.6
Scotland	19.05	14.64	5256	3.5	5.4
Sweden	15.70	10.81	2553	3.4	4.9
U.S.A.	16.15	13.34	6231	3.5	6.8

[a] See note concerning Finnish data on page 6.

as the standard deviation of age. The standard deviation of age for Population 1b is really a better index of the amount of grade repeating practised, since Population 1a is a chronological population taken from across grades. Thus, the standard deviation of age for Population 1b is repeated in this table. (Table A.7 in the Appendix presents for Population 1a the means, standard deviations and number of students for each country by sub-sample for Total Mathematics Score, Higher Mental Process and Lower Mental Process.).

The spread of mathematics scores in Japan, Sweden and the United States is much the same as for Population 1b. England and Scotland have small standard deviations and Australia, Belgium, Finland, France and the Netherlands have larger standard deviations. Although this indicates that where an age group is spread over grades its standard deviation is larger than when a grade group is spread over ages (again because the teacher is teaching to a grade level), it is still interesting to note that England (inter- and intra-school differentiation) and Japan (severe *de facto* inter-school differentiation) have the largest standard deviations. However, it is to be expected that the chronological population's (1 a) standard deviation will be more strongly associated with the index of grade repeating that has been chosen than the standard deviation of the grade population (1 b). The correlation is −.05, which although negative is less so than the correlation of −.53 in Population 1 b between the mathematics score and grade repeating. This supports the theory that the standard deviation will be larger where an age group is spread across grades than when a grade group has some other ages in it. However, before arriving at any overall conclusions let us also examine these standard deviations in conjunction with measures of the amount of ability grouping practised in each of the systems.

Intra-School Differentiation—Ability Grouping

Each school principal of the schools in the sample was asked to respond to the following question on the School Questionnaire:

To what extent does educational differentiation (e.g. setting, streaming, ability grouping) take place within your school?

It is universally practised	1
It is generally practised	2
It is practised in some age or grade groups only	3
It is practised at all	4
Comment	

This was asked in various ways in the various countries, but always coded according to the above *international* frame of the question. The United States phrased their question as follows:

To what extent does ability grouping take place within your school?

It is practised for all pupils	1
It is practised for some pupils at all levels	2
It is practised in some age or grade groups only (Indicate in which groups under "Comment")	3
It is not practiced at all	4
Comment	

the French as follows:

Dans quelle mesure pratiquez-vous la sélection : (entourez le numéro correspondant)

Toujours	1
Généralement	2
Relatif à un certain âge ou à un certain niveau	3
Jamais	4

Donnez les raisons de votre action

and the English as follows:

To what extent does educational differentiation (e.g. setting, streaming, ability grouping) take place in your school?

It is universally practised	1
It is generally practised	2
It is practised in some age groups only	3
It is not practised at all	4
It is practised in mathematics at all ages	5
It is practised in mathematics in some age groups only	6
It is practised in one or more other subjects at all ages	7
It is practised in one or more other subjects in some age groups only	8
Comment	

Two indices were derived from the data. The first was a mean score based on the code 1–4 where a low number devotes ability grouping is practised a great deal and a high number means it is practised little or not at all. The second was the percentage of all school principals responding to either the first statement (universal) or the second (general). Table 6.3 presents these data for both Populations 1a and 1b.

Since the first index is based on all of the responses and not just two as in the case of the second index, it is the first index which will be used. There is, of course, a very close similarity in the ranks. Some

Table 6.3. *Indices of the extent of ability grouping practised.*

Country	Population 1a Ability grouping (1)	Population 1a Ability grouping (2)	Number of schools	Population 1b Ability grouping (1)	Population 1b Ability grouping (2)	Number of schools
Australia	2.63	48	108	2.63	48	72
Belgium	2.47	54	61	2.47	57	61
England	2.12	64	184	2.13	64	182
Fed. Rep. of Germany	—	—	—	3.83	0	161
Finland	4.0	0	111	4.0	0	111
France	3.0	45	125	3.02	20	124
Israel	—	—	—	3.44	2	154
Japan	3.88	0	210	3.88	0	210
Netherlands	3.14	9	88	3.11	10	30
Scotland	1.75	77	73	1.73	78	73
Sweden	2.69	36	80	2.69	34	80
U.S.A.	2.19	62	395	2.21	66	395

comment on these indices seems appropriate at this point; it is difficult to believe that in Finland there is no ability grouping whatsoever, especially since some inter-school differentiation is practised. The United States schools seem to practise ability grouping much more than one would have expected. Although there may have been some error in filling in the responses, it is, however, unlikely to have been consistent when one observes the number of the schools involved.

Sweden ranks sixth in the amount of ability grouping practised in the schools as a whole, although it must be remembered that no intra-school differentiation officially took place until seventh grade. The product-moment correlations between the extent to which ability grouping is practised in a system and the standard deviation of Total Mathematics Score is −.29 and −.18 (the negative sign is a result of the code) for Populations 1 b and 1 a respectively. This supports the hypothesis that by forming homogeneous groups of ability or achievement within an overall age or grade group, the overall group will become more heterogeneous in its achievement than if it were taught without differentiation. It is clear that the greater the extent to which ability grouping is practised, the wider are the standard deviations of scores. However, it is also important to examine the relationship between ability grouping and the standard deviation of mathematics score when grade repeating and the mean mathematics score are held constant.

Table 6.4 presents for Population 1a the standard deviations of total mathematics scores for each country as well as the measure of ability grouping, grade repeating and mean mathematics score. The latter is included since it has already been noted that there is a substantial correlation between mean score and standard deviation.

Tables 6.5 and 6.6 present the product-moment correlation matrix of Table 6.4 and the simple correlations and regression weights[3] of ability grouping, grade repeating and mean mathematics score as predictors with the criterion (standard deviation). The third column of Table 6.6 gives the contribution to the total variance (multiplied by 100) of each of the predictors.

It is evident that ability grouping is strongly associated with large

[3] The multiple regression procedure used was that reported by Cooley and Lohnes, *Multivariate Procedures for the Behavioral Sciences*, Wiley, New York, 1962, pp. 31–59.

Table 6.4. *Standard deviations, measures of ability grouping and grade repeating and mean mathematics scores.*

(Population 1a)

Country	s.d. of math. scores (1)	Ability grouping (2)	Grade repeating (3)	Mean score math. (4)
Australia	14.01	2.63	7.70	20.18
Belgium	15.02	2.47	8.80	27.74
England	16.97	2.12	4.20	19.31
Finland	10.76	4.00	6.66	15.39
France	12.37	3.00	7.80	18.32
Japan	16.90	3.88	3.40	31.16
Netherlands	15.91	3.14	11.60	23.86
Scotland	14.64	1.75	5.40	19.05
Sweden	10.81	2.69	4.90	15.70
U.S.A.	13.34	2.19	6.80	16.15
Grand Mean	14.07	2.79	6.73	20.69
Grand s.d.	2.26	0.73	2.42	4.79

Table 6.5. *Product moment correlation matrix of Table 6.4.*

	1	2	3	4
1	1.000	−.181	−.047	.726
2		1.000	.039	.265
3			1.000	.060
4				1.000

Table 6.6. *r, b and rb 100 of Table 6.5.*

	r	b	rb 100	
Ability grouping	−.181	−.399	7.22	$R^2 = 0.684$
Grade repeating	−.047	−.082	0.38	$R = 0.827$
Mean TMS (corr.)	.726	.837	60.77	
Total variance accounted for			68.37	

standard deviations in both populations (the negative signs are mere consequences of the coding used). As expected, grade repeating is associated with small standard deviations in Population 1b (the grade population) but has practically no association with the size of the standard deviation in Population 1a (the chronological popula-

Table 6.7. *Standard deviations, measures of ability grouping and grade repeating and mean mathematics scores.*

(Population 1b)

Country	s.d. of math. scores (1)	Ability grouping (2)	Grade repeating (3)	Mean score math. (4)
Australia	12.28	2.63	7.70	18.88
Belgium	13.75	2.47	8.80	30.43
England	18.53	2.13	4.20	23.76
Fed. Rep. of Germany	11.70	3.83	6.60	25.45
Finland	11.61	4.00	6.66	16.13
France	13.23	3.02	7.80	20.96
Israel	14.67	3.44	5.60	32.29
Japan	16.90	3.88	3.40	31.16
Netherlands	12.12	3.11	11.60	21.43
Scotland	15.69	1.73	5.40	22.31
Sweden	10.83	2.69	4.90	15.26
U.S.A.	13.21	2.21	6.80	17.85
Grand Mean	13.71	2.93	6.62	22.99
Grand s.d.	2.34	0.75	2.21	5.82

Table 6.8. *Product-moment correlation matrix of Table 6.7.*

	1	2	3	4
1	1.000	−.294	−.535	.544
2		1.000	.011	.220
3			1.000	−.164
4				1.000

tion, where students of the same age are spread across several grades). Again, as would be expected, the mean score contributes considerably to the variance since it was known that the distribution of the scores on the tests tended to be crowded towards the foot and open at the top.

From other researches already mentioned at the beginning of this chapter, there is evidence concerning the effect of grouping practices on lower socio-economic groups in some systems of education, but before proceeding to consider some of the implications of the results

Table 6.9. *r, b and rb 100 of Table 6.8.*

	r	b	rb 100	
Ability grouping	−.294	−.422	12.41	$R^2 = 0.670$
Grade repeating	−.535	−.448	23.97	R = 0.819
Mean TMS (corr.)	.544	.563	30.63	
Total variance accounted for			67.01	

presented in this chapter, it is well to reflect on certain limitations to the findings. First, there is no separation of setting from ability grouping in the measure of ability grouping—thus the measure is impure. The measure of grade repeating is an inferred measure. Purer measures should in future be obtained. With a maximum of 12 observations (in this case countries) a multiple regression analysis containing more than three predictors is inadvisable because of the few remaining degrees of freedom. If we had more systems in the analysis—either more countries or sub-divisions of countries—this analysis could be pushed much further.

Implications

What are the educational implications of these findings? Some European countries are considering changing from a selective school system to a comprehensive system (e.g. England). Sweden has already done so and about half of Scotland's secondary schools are comprehensive. It should be realised by policy makers that to eliminate inter-school differentiation but to retain intraschool differentiation (ability grouping) will still mean a fairly large variability of achievement, although perhaps not quite so large as before. The principle of ability grouping within schools is exactly the same as that of inter-school differentiation. Many teachers (Yates, 1966) believe in ability grouping and even though teachers or head teachers are in a deliberate non-ability grouping school they will occasionally indulge in it subconsciously—for example, the head teacher who says: "Ah, yes, I have no streaming in my school; in this class X, for example, there are pupils of very different ability, an absolutely heterogeneous group: the bright ones are over there on the right hand side, the not so bright in the middle, and the poor ones on the left." In other

words, it is the philosophy of the teachers which it is important to change; it would be insufficient to take an administrative decision that there should be no more ability grouping in schools without also helping the teachers to change their outlook. This may be particularly difficult in countries such as England and Scotland, where the capacity theory of intelligence is very prevalent, not only among teachers, but also among some educational policy makers (Pidgeon, 1966).

There is evidence (Svensson, 1962, and Husén, 1966) to indicate that "good" students are not held back by "poor" students when in the same school and, what is more important, that "poor" students improve when with "good" students, whereas when put into a homogeneous group they deteriorate. Thus, where differentiation is being practised at an early stage in the school system, it is the "culturally-disadvantaged" and/or lower ability child who suffers. In a sense, the practice of differentiation can exacerbate the plight of the culturally-disadvantaged child, since once differentiated into the "poor" ability group (either inter- or intra-school) he will, in relation to his peers (age group) deteriorate—wide standard deviations —rather than improve—narrow standard deviations (cf. Robbins Report, Appendix I).

The evidence provided in this chapter is based on differences between educational systems, and it would seem that administrative decisions concerning both inter- and intra-school differentiation can affect the size of the standard deviation in mathematics scores. Whether the same would hold true in other subject areas is a matter for future research, but it would seem likely. Educational policy makers should be aware of these facts when considering any changes in their school systems.

Summary

The relationships between three aspects of differentiation and the variability of mathematics scores on the IEA tests are examined in the light of data from twelve different systems of education. The three aspects are (1) inter-school differentiation, (2) intra-school differentiation (grade repeating) and (3) intra-school differentiation (ability grouping). After a discussion of relevant previous research, both at the international and national levels, an examination was made of the standard deviations for Populations 1b and 1a. Population 1b was chosen as the main focus of attention, since it is the last grade still in compulsory schooling in all of the countries in the study. Interpretation of the size of the standard deviations in each country was undertaken in terms of the three aspects of differentiation mentioned. Unfortunately, no suitable index of inter-school differentiation exists, but it would seem that either *de facto* or *de jure* inter-school differentiation does tend to be associated with wide standard deviations.

The standard deviation of age of Population 1b was used as a measure of grade repeating, and it was found in Population 1b that the greater the degree of grade repeating the smaller the standard deviation. However, the association in Population 1a was, as expected, nearly zero.

Specific data were collected from the school principals of schools in the sample on the extent to which ability grouping was practised in their schools as a whole. The mean score on this variable for all schools in the target population within a country served as the index of the extent to which ability grouping was practised in each country. There was a correlation of about .25 between the size of standard deviations and the extent to which ability grouping was practised. When grade repeating was partialled out of the correlation between standard deviation and ability grouping the correlation was about .4. When ability grouping was partialled out of the correlation between standard deviation and grade repeating, the 1b Population correlation became about –.4 and the 1a Population remained near zero. This indicated that grade repeating was associated with a lower standard deviation for Population 1b while for Population 1a there was no association.

Differentiation into homogeneous groups (inter-school differentiation and intra-school differentiation—ability grouping) within age groups was found to be associated with large standard deviations. Grading and grade repeating is associated with small standard deviations. Educational policy makers should be aware of the relationship between these educational practices and the spread of scores on achievement tests in mathematics. This is of particular importance in the debate concerning selective versus comprehensive systems of education. Ability grouping within schools is associated with large standard deviations in a school system, even though that school system may have no inter-school differentiation. Furthermore, it is not enough to take an administrative decision concerning differentiation without, at the same time, changing teachers' attitudes about differentiation. These findings are also of interest to those concerned with the "culturally disadvantaged" child, since certain differentiation practices can exacerbate his plight, whereas it would appear that non-differentiation might improve it. It must be remembered that these findings are concerned with one subject area only, and must be checked by future research in other subject areas.

CHAPTER 7

Specialization and Age of Entry to School

Two separate aspects of school organization are examined in this chapter. The first concerns the relationship between countries of the number of subjects studied (specialization) in the pre-university year by the mathematics group (Population 3a) and the mean mathematics achievement score. The second concerns the relationship between countries of the 13-year-olds (Population 1a) and 13-year-old grade (Population 1b) mean mathematics scores and the age at which they entered school.

Specialization

In some educational systems, pre-university year students study only three or four subjects and have been doing so since the age of sixteen (England and Scotland) whereas in other countries all students are expected to continue studying nine or more subjects to the end of their secondary school career. The English position is based on the alleged virtues of study in depth. The Swedish position (9 subjects) may be based partly on the assumption that, given the rapidity of technological change, which means that many of the next generation will almost inevitably have to be occupationally retrained at least more than once in their working lives, it seems that a broader education is more appropriate for the academically gifted.

In those countries where specialization occurs, it often happens that students begin dropping subjects as early as 13 years of age (e.g. England—see Jackson, 1966) and by the age of 16, there is an evident bias (arts versus science subjects) in the cluster of subjects studied. Does specialization really lead to a greater knowledge of the subject studied? It is possible to examine this in the light of the IEA data—knowledge in this case being defined as the mean achievement scores on the IEA mathematics tests. In Population 3a, all students were studying mathematics in the pre-university year; in some educational systems, however, mathematics was studied in conjunction with only two other subjects, whereas in other systems, it was studied in conjunction with eight or more.

There has been a great deal of discussion about the values of specialization, but no research appears to have been carried out. This is perhaps not surprising, since within systems of education there has been a uniformity of practice. Furthermore, where a system has had students specializing in three or four subjects only, it has been the brighter students who have studied four or five subjects and who would therefore be likely to be higher scorers than those only studying two or three subjects. The IEA Study is the first large-scale international study of its kind, and therefore this is the first time that comparisons can be made of achievement between groups studying a limited number of subjects and those studying more.

In the School Questionnaire, a question was asked about the average number of subjects studied in each grade in the school. Unfortunately, the data obtained are limited in application, since in some countries different interpretations have been put on the word "subject" by different head teachers. Some have interpreted all "subjects" as including sport and drama, whereas others have included academic subjects only. However, the data given in the Case Study Questionnaire on the "number of subjects studied" would appear to be in order. Table 7.1 indicates the average number of subjects studied per country (according to the Case Study Questionnaire), the mean corrected mathematics score in each country, and the standard deviation and the number of students.

If the eight countries showing eight or more subjects studied are combined to form one group, and the three countries showing four or fewer subjects are combined to form a second group, then the mean scores of the two groups are found to be 31.1 and 24.8 respectively, giving a difference of 6.3, which is highly significant. Students from countries where 8 or more subjects (of which mathematics is one) are studied at the pre-university level perform better in mathematics than students from countries where only four or less subjects (of which mathematics is one) are studied. This is contrary to expectation.

There are, however, complications. The United States system is

Table 7.1. *Number of subjects studied and mean score by country.* (Population 3a)

Country	No. of subjects studied	Mean score	Standard deviation	Number of students
Belgium	9 +	34.6	12.6	519
France	9 +	33.4	10.8	222
Netherlands	9 +	31.9	8.1	462
Japan	9 +	31.4	14.7	818
Finland	9	25.3	9.6	369
Fed. Rep. of Germany	9	28.8	9.8	649
Sweden	9	27.3	11.9	776
Israel	8	36.4	8.6	146
Australia	6	21.6	7.5	1089
Scotland	4	25.5	10.5	1422
U.S.A.	4	13.8	12.6	1568
England	3	35.2	12.6	967

not as specialized as it would appear from the entry in the table, because although it may be the case that only four "solids" are studied in 12th grade, they may not be the same "solids" as in 11th grade (or perhaps only one or two are the same in both grades) and thus the actual number of subjects studied in the last two grades could range from four to seven or eight. If the United States is omitted from the specialist group, the average score of that group then becomes 30.4, which is not significantly different from the average of 31.1 of the first group.

Assuming the IEA mathematics tests to be fair tests of mathematical achievement of the various pre-university populations studied, it is surprising that students from specialization countries do not score significantly higher than students from non-specialization countries. It should be pointed out perhaps that some English mathematical educators have stated that they did not think that the IEA mathematics test extended the best students. Furthermore, in England, the syllabus in Applied Mathematics was covered only to a very small extent by the tests. Because of the wide range of scores between countries within each of the two groups, it would seem that there are obviously factors other than the number of subjects studied which account for the differences.

The average ages of the students in the eight countries (i.e. the first eight countries in Table 7.1) are, with one exception (Japan), over 18, while the average ages of the students in the four remaining countries are all under 18. Taking more subjects thus appears to be associated with a higher age, the assumption being that students must prolong their school education to be able to carry the extra load.

It is perhaps also of interest to note that in two of the three countries in which four or fewer subjects are studied there is a mandatory age of entry to school of five years. The question of differing degrees of retentivity has been dealt with in Chapter 5, but is also relevant in these comparisons. It is striking that the students in Israel and Belgium do not differ very much from English students in age, since Belgium has approximately the same degree of retentivity as England and the mean mathematics scores of each of these countries are close to each other, even though in England the average number of subjects studied is five less than in the other countries.

The conclusion 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, will probably be of interest to educational policy makers and planners in England, Scotland and Australia. However, it must be emphasised that this study of specialization is extremely limited because of the wide differences on several important independent variables which have not been held constant in this analysis. It is important that further work is carried out both nationally (cf. Pidgeon *et al.*, 1967) and internationally. Hopefully, with IEA continuing in six subject areas it will be possible to examine the effects on other subject areas when specialization takes place in a particular subject.

Age of Entry to School

In each country there are regulations specifying when "normal" children (i.e. excluding such children as spastics, extremely mentally retarded, etc.) should at the latest begin compulsory schooling. In some countries (e.g. Sweden and Germany) there is a single day in the school year on which all children within a year age range begin school. In others (e.g. Scotland and England) there are two or three possible days of entry. In most areas in England, for example, all

8 – 671266 *Postlethwaite*

children who will be five years of age between September and the end of December begin school on the first of September; those who will be five between January and the end of March begin on the first of January; and those who will be five between April and August begin in the middle of April.

As with most general regulations, there are exceptions. In certain countries children slightly younger than the mandatory age of entry may begin school if there are exceptional grounds. It is usually the local school authority which then decides whether or not the grounds are exceptional. In several European countries it is possible for children to start school before they reach the mandatory age, if they can prove that they are "mature" enough for school. The judgement of this maturity has, up to the present time, involved physical tests of fitness for school, as well as certain group tests of reasoning. Examples of this testing are the *skolmognadstest* in Sweden and the *Schulreife* test in Germany.

It should be remembered, furthermore, that in all countries preschools are attended in different degrees. For example, in some English-speaking countries there are nursery schools and kindergartens, but it is only a small percentage of an age group which attends. In the United States, however, about fifty percent of children attend kindergarten. In the French-speaking countries it is estimated that approximately 50 percent of an age group attend the *école maternelle* (or *jardin d'enfants*). Thus, the differences in amounts of preschooling must be borne in mind when comparing at a later stage the performance of students from countries with different mandatory ages of entry of school.

As far as previous research is concerned, there are two cross-country studies which have examined, in part, the effect of differing amounts of formal schooling to which children in different countries have been exposed. Anderson (1964) has suggested that the superiority of the performance of English and Scottish children over American students at the age of seven can be attributed to the extra year of schooling. But when differences occurred at ages ten and fourteen, he preferred to explain these in terms of differences in instruction. Similarly, Pidgeon (1958), although finding English 11-year-old children superior to 11-year-old California children (English mean = 29.1, standard deviation = 18.7 and California mean = 12.1, standard deviation = 6.8 on a 70 item test), states that the main reasons for the different levels in performance are probably due to the fact that formal teaching tends to be introduced at an earlier age in England, and to the fact that there is a difference in the standards in the two systems. He points out that in the United States more limited objectives are formulated for children of primary school age and less emphasis is placed on progress in mechanical arithmetic than is customary in England.

A national study which has relevance to this problem was carried out by Mogstad (1958) in Norway. It occurred that 12-year-old students in a rural region of Norway were in two parallel groups. One group received the full week regular schooling for two years. The second group received formal schooling for only half this period (i.e., half the amount of formal instruction), although it must be noted that the second group undertook much more homework due to the fact that they were in sparsely populated areas and could attend school for only half the time. In specially constructed achievement tests, the second group was only slightly inferior in performance at the end of the two years to the first group, even though the number of periods devoted to each subject was half.

The IEA Study is the first study undertaken where it has been possible to examine differences between the performance of fully representative samples from more than three countries in a particular school subject. Here, it has been possible to compare the performances of 13-year-olds in countries having mandatory ages of entry to school at five (two countries), six (six countries), or seven (two countries).

The two populations which it is relevant to examine in connection with this problem are the 13-year-olds in each system (i.e., the 1a Population) and students in the grade where most 13-year-olds are to be found (i.e., the 1b Population).

The 1a Populations are chronologically comparable and are directly related by age to the mandatory age of entry to school. If the various lengths of schooling up to the age of thirteen years make a difference, then it should be apparent in this analysis.

The second population is the grade population in which most 13-year-olds are to be found. Two extra countries to those in Population 1a are represented in Population 1b and for this reasons the 1b results are also presented. The actual grades tested have been given in Chapter 4. Although the standard deviations of age for Populations

Table 7.2. *Mean ages and standard deviations of age for populations 1a and 1b.*

Country	Population 1a: Mean age in months	Population 1a: Standard deviation	Population 1b: Mean age in months	Population 1b: Standard deviation
Australia	161	3.5	159	7.7
Belgium	162	3.3	168	8.8
England	162	3.3	172	4.2
Fed. Rep. of Germany	—	—	164	6.6
Finland[1]	163	3.3	167	6.7
France	162	3.5	163	7.8
Israel	—	—	167	5.6
Japan	161	3.4	161	5.4
Netherlands	163	3.1	157	11.6
Scotland	160	3.5	168	3.4
Sweden	163	3.4	164	4.9
U.S.A.	163	3.5	164	6.8
Median	162	3.4	164	6.7
Range	3	0.4	15	8.2

[1] See note on Finnish data on page 6.

1a and 1b have already been given in Chapter 6, they are repeated here in Table 7.2 together with the mean ages of these populations in each country.

In Table 7.3, countries are grouped into three groups according to whether the mandatory age of entry is five, six, or seven years of age. The median age of entry for each country is given. The source for these figures is the National Case Study Questionnaire. Table 7.3 also gives the means, standard deviations and number of students for the various groups. The averages for the different groups of countries are simple and not weighted averages. If averages were weighted according to the number of students tested in each country, they would be biassed towards the averages of those countries where most students were tested. This is not what is required, but straight averages with each country regarded as a single observation.

It is interesting to note that although the regulations for entry to school in England and Scotland differ, the actual median age of entry is the same. In England, the regulation is that children who will become five years of age up to and including the first day of

Table 7.3. *Mean scores and standard deviations of scores in mathematics for different ages of entry.*

Country	Mandatory age of entry	Median age of entry	Population 1a: M	Population 1a: S.D.	Population 1a: N	Population 1b: M	Population 1b: S.D.	Population 1b: N
England	5 yrs	5 yrs 2 mo.	19.2	17.0	3012	23.7	18.5	3148
Scotland	5 yrs	5 yrs 2 mo.	19.1	14.6	5256	22.3	15.7	5718
			19.2			23.0		
Australia	6 yrs	5 yrs 7 mo.	20.2	14.0	2916	18.9	12.3	3078
Belgium	6 yrs	6 yrs 2 mo.	27.7	15.0	1686	30.4	13.7	2644
Fed. Rep. of Germany	6 yrs	6 yrs 5 mo.	—	—	—	25.5	11.6	4476
France	6 yrs	6 yrs 0 mo.	18.3	12.4	2410	21.0	13.2	3549
Israel	6 yrs	6 yrs 0 mo.	—	—	—	32.3	14.7	3232
Japan	6 yrs	6 yrs 0 mo.	31.2	16.9	3049	31.2	16.9	2049
Netherlands	6 yrs	6 yrs 5 mo.	23.9	15.9	428	21.4	12.1	1444
U.S.A.	6 yrs	6 yrs 5 mo.	16.9	12.7	6231	17.9	13.3	6544
			23.0			24.8		
Finland[2]	7 yrs	6 yrs 8 mo.	15.4	10.8	1156	16.1	11.6	1325
Sweden	7 yrs	7 yrs 0 mo.	15.7	10.8	2553	15.3	10.8	2828
			15.6			15.7		

[2] See note on Finnish data on page 6. In Table 7.3 the scores given for Finland are the corrected scores.

next term begin school on the first day of this term. In Scotland, it is those children who have become five years of age since the beginning of last term who begin school the first day of this term. Thus, one would expect the median age of entry to be about 4 years 10 months in England, and 5 years 2 months in Scotland. However, it would appear that because of a shortage of places in Infants Schools in England, there is a delay in children's entering school.

The differences in means are listed in Table 7.4.

Table 7.4. *Differences between mean scores of groups with different ages of entry.* (Populations 1a and 1b).

Population	6 yrs v. 5 yrs	7 yrs v. 6 yrs	7 yrs v. 5 yrs
1a	3.8	−7.4	−3.6
1b	1.8	−9.1	−7.3

The application of the test of the difference being more than twice the complex standard error of sampling indicates that all the differences are statistically significant and that countries with an entrance age of six produced, on average, higher scores than those where children enter school at 5 or 7 years of age. There is little difference between the two countries with a 5 year entry; a weak majority of countries with a 6 year entry do better than these two, but the two countries with a 7 year entry do worse. This suggests that some loss attends delaying the entry until 7 years.

Age of Entry and Social-Status Groups

It was possible to break down the scores for Population 1a by social status groups. Table 7.5 presents the scores for social groups 1–6 and for groups 7, 8 and 9 separately. The definitions of these social groups are given in full in Volume I, Chapter 8 of Husén *et al.* (1967). The following is a brief description of each:

Group 1—Higher Professional and Technical
Group 2—Administrators, Executives and Working Proprietors: *large and medium scale*
Group 3—Sub-Professional; Technical
Group 4—Small Working Proprietors (other than in agriculture, forestry, or fishing)
Group 5—Proprietors and Managers in Agriculture, Forestry, Fishing
Group 6—Clerical and Sales Workers (lower levels of white collar work)
Group 7—Manual Workers: Skilled and Semi-Skilled
Group 8—Labourers (hired) in Agriculture, Forestry, Fishing
Group 9—Unskilled Manual Workers (excluding agriculture, forestry, fishing)
Group 0—Unclassified; No Answer

Although it would appear that children from social groups 1 to 6 (professional and white-collar workers) benefit more from early entry to school than do children from groups 7 to 9 (farmers and blue-collar workers), it is difficult to draw firm conclusions because of the heterogeneity of scores within each of the age entry groups. There are some interesting differences between social groups within coun-

Table 7.5. *Mean score in mathematics by social-status group.*
(Population 1a)

	Groups 1–6			Group 7			Group 8			Group 9		
Country	M	S.D.	N	M	S.D.	N	M	S.D.	N	M	S.D.	N
England	29.54	17.19	931	15.50	14.69	1764	16.09	11.66	50	27.61	17.32	10
Scotland	26.33	14.88	1456	17.13	13.57	3180	17.04	13.77	122	13.27	12.54	171
Total	27.90	16.03	2387	16.81	14.13	4944	16.56	12.71	172	20.44	14.93	181
Belgium	31.62	14.17	863	24.83	14.72	662	24.49	21.99	9	21.19	13.92	107
France	21.88	13.31	895	16.85	11.09	1249	15.27	10.59	39	13.82	12.05	80
Netherl.	29.47	16.21	210	19.28	13.70	185	14.64	9.97	20	21.01	18.24	8
Japan	33.30	16.61	1406	28.05	16.27	485	23.07	14.87	45	21.68	17.52	24
U.S.A.	20.17	13.62	2916	13.89	12.06	2645	12.23	10.45	102	12.89	11.78	28
Australia	23.68	13.93	1380	18.15	13.18	1219	13.55	12.80	79	14.34	10.79	110
Total	26.69	14.64	7670	20.10	13.50	6445	17.21	13.44	294	17.49	14.05	357
Finland*	23.87	9.53	407	24.17	10.12	301	18.17	11.33	9	17.19	9.79	25
Sweden	17.62	11.13	1226	14.45	10.15	1075	11.42	7.92	99	12.21	8.33	49
Total	20.74	10.33	1633	19.41	10.13	1376	15.07	9.62	118	14.70	9.06	74

* The data here are the uncorrected Finnish data. It has not been possible to rerun these data since the mistake in the Finnish data was discovered.

tries in Table 7.5. Group 7 in Finland has a higher score than Groups 1 to 6.[a] The direction of the scores in Groups 7 to 9 in England is contrary to expectation (although the differences are not statistically significant).

The actual differences in scores from Table 7.5 are reported in Table 7.6.

Table 7.6. *Differences between mean scores in Table 7.5.*
(Population 1a)

	Groups 1–6	Group 7	Group 8	Group 9	Groups 8 and 9 combined
5 yrs v. 6 yrs	1.21	**−3.79**	−0.65	2.95	1.19
5 yrs v. 7 yrs	**7.16**	**−3.10**	1.49	5.74	3.62
6 yrs v. 7 yrs	**5.95**	0.69	2.14	2.79	2.43

[a] This is more likely to be a result of incorrect weighting than a realistic fact—see note on Finnish data on page 6.

It is clear that to make the mandatory age of entry to school earlier (e.g. from 6 to 5) 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 bluecollar workers. It is the qualititative differences which must now be the subject of more systematic research.

Further Analyses Related to Age of Entry

It has been pointed out in Chapter 3 in Volume II of Husén *et al.* (1967) that "when the 3a Population scores are adjusted for differences in the proportions of an age group still at school, it is found that the gains between 1a and 3a stages are directly related to the time interval between the two stages, the rate of gain being the same in practically all of the countries". In other words, the differences in scores between countries are already established by the age of 13.

Since this book is concerned with organizational aspects of educational systems, it is worthwhile examining the relationship of certain other organizational features in addition to age of entry to school to the differences in mathematics scores between countries of 13-year-old students.

The number of subjects studied in grade 8 (the grade where most 13-year-olds were to be found) is of interest. Is, for example, the studying of fewer subjects associated with higher scores at this level? The number of subjects on average studied in each school was collected by means of the School Questionnaire. The figure given in Table 7.7 is the average for each country. There is considerable difference in the length of preservice training of teachers as between countries; this information consisting of the number of post-secondary school years preservice training was collected from the Teacher Questionnaire. Within countries, interest in mathematics accounts for a considerable amount of the variance and it is therefore of interest in a between countries analysis. The interest score was derived from various pieces of information collected in the Student Questionnaire. The higher the score the greater the interest. (The derivation of this index is explained in detail on pages 212–213 in Volume I of Husén *et al.*, 1967). There is also considerable variation between countries on the number of hours a week spent both in school and on homework. These data were collected through the Student

Table 7.7. *Mean mathematics score and measures of various independent variables.* (Population 1a)

Country	Total math. score (1)	Age of entry (2)	No. of subjects grade 8 (3)	Pre-service training (4)	Inter. in math. (5)	Hrs school per wk in math. (6)	Hrs home-work per wk in math. (7)
Australia	20.18	6.0	8.7	2.8	59	38	24
Belgium	27.74	6.0	8.9	2.4	57	62	36
England	19.31	5.0	8.9	3.1	57	38	17
Finland	15.39	7.0	9.0	3.2	58	30	24
France	18.32	6.0	8.5	2.1	55	45	34
Japan	31.16	6.0	9.0	3.2	61	39	30
Netherlands	23.86	6.0	9.0	4.1	54	44	26
Scotland	19.05	5.0	8.2	4.0	53	43	23
Sweden	15.70	7.0	9.0	4.6	58	57	19
U.S.A.	16.15	6.0	7.3	4.4	62	47	31
Grand mean	20.69	6.0	8.65	3.39	57.40	44.30	26.40
Grand s.d.	5.31	0.67	.54	0.85	2.88	9.40	6.24

Questionnaire and again the higher the number the greater the number of hours.[4]

Table 7.7 presents the data on each of the above variables as well as on Mean Mathematics Score for Population 1a. For convenience, columns 5, 6 and 7 have been multiplied by ten. The data are presented only for those countries for which data on all of these variables are available.

Table 7.8 presents the product-moment correlation matrix from Table 7.7. Table 7.9 presents the simple correlations, regression coefficients and their products multiplied by 100.

With as many as six constants fitted to ten observations it is clear that the multiple correlation will be rather spuriously high. None the less the regression coefficients are perhaps worth some attention. Let us take them in turn. The large negative coefficient for "age of entry" reflects chiefly the fact (see Table 7.3) that the countries delaying age of entry until the age of seven are low scorers. The large

[4] For detailed information on how the data in this paragraph were collected (except for "Interest in Mathematics") see Appendix II of Volume I of Husén *et al.*, 1967).

Table 7.8. *Product-moment correlation matrix of Table 7.7.*

	1	2	3	4	5	6	7
1	1.000	−.228	.377	−.334	.070	.180	.452
2		1.000	.276	.137	.348	.106	.080
3			1.000	−.301	−.249	−.068	−.268
4				1.000	.097	.110	−.462
5					1.000	−.071	.139
6						1.000	.356
7							1.000

Table 7.9. *r, b and rb 100 of Table 7.8.*

	r	b	rb 100	
Age of entry	−.228	−.894	20.38	$R^2 = 0.967$
No. of subjects in grade 8	.377	1.208	45.54	$R = 0.983$
Pre-service training	−.334	.622	−20.77	
Interest in math.	.070	.465	3.25	
Hours school per week in math.	.180	−.067	−1.21	
Hours homework per week in math.	.452	1.096	49.54	
Total variance accounted for			96.73	

positive coefficient for "number of subjects in Grade 8" reflects chiefly the fact that the United States is a low scoring country. It is only in the United States that the number of subjects differs by more than 1 from the general average. This is the analytical explanation of this high coefficient but it is hard to believe that this fact in itself can be a main part of the reason why the United States is a low scorer; it seems much more likely that this is not a case where A is the cause of B or vice-versa, but rather a case where A and B are both caused by something else.

The high coefficient for "pre-service training" is on a different footing; common sense suggests that there may well be a causal relation here. "Interest in Mathematics" has a high coefficient, which may well correspond to a causal relation, though the direction is less clear. Does interest in mathematics promote good performance, or good performance promote interest in mathematics? It is possible for observers to hold different views on this. The remaining high coefficient is for "hours of homework per week" and this strongly suggests a causal relation.

One important independent variable within countries proved to be the teachers' rating of the student's opportunity to learn the items in the test (see Husén *et al.*, 1967). Each teacher was asked to rate on a three point scale the proportion of his students taking the test having had the opportunity to learn each item.[5] These data were then averaged percentwise for each country. Table 7.10 presents these data for the eight countries where they were available as well as repeating in addition the measures of pre-service training, interest and hours school per week which have already been used above.

The most striking feature of Table 7.12 is the large contribution made by "Opportunity to learn". What can this mean? The face meaning is clear enough. In the low scoring countries fewer boys and girls had covered the subject matter of the tests. Can the reason

Table 7.10. *Mean mathematics score and measures of pre-service training, opportunity to learn, interest and hours school per week in Mathematics.*

(Population 1a)

Country	Total math. score (1)	Pre-service training (2)	Opp. to learn (3)	Interest in math. (4)	Hours school per week in math. (5)
England	19.31	3.1	60	57	38
Finland	15.39	3.2	47	58	30
France	18.32	2.1	50	55	45
Japan	31.16	3.2	63	61	39
Netherlands	23.86	4.1	52	54	44
Scotland	19.05	4.0	51	53	43
Sweden	15.70	4.6	37	58	57
U.S.A.	16.15	4.4	48	62	47
Grand mean	19.90	3.59	51.00	57.25	42.87
Grand S.D.	5.31	.83	8.00	3.20	7.81

[5] For further details see Chapter 4 of Volume II of Husén *et al.*, 1967.

Table 7.11. *Product-moment correlation matrix of Table 7.10.*

	1	2	3	4	5
1	1.000	−.176	.751	.093	−.194
2		1.000	−.473	.157	.526
3			1.000	.078	−.572
4				1.000	−.010
5					1.000

be merely that the choice of subject matter of the tests was unsuitable for these countries and that they might have done better had there been a different choice of subject matter. On the whole this seems unlikely. It is certainly less likely at this level than at the higher level (Population 3a). At the higher level there is much difference of opinion, both within countries and between countries, about what the mathematical curriculum ought to be but about the curriculum at the age of thirteen there is a fairly close consensus. It seems likely therefore that in countries where the index for "Opportunity to learn" was low the students have made less progress in covering a broadly international curriculum than those in countries where the index was high. The countries where the index is low are the countries where compulsory schooling extends longer. They are in fact the United States, Sweden and Finland. In the two Scandinavian countries compulsory schooling does not begin until seven. In the United States the proportion staying on after the compulsory stage of schooling is high. A late entry would account for the fact

Table 7.12. *r, b and rb 100 of Table 7.11.*

	r	b	rb 100	
Pre-service training	−.176	.132	−2.32	$R^2 = 0.658$
Opp. to learn	.751	.984	73.90	$R = 0.811$
Interest in math.	.093	—	—	
Hours school per week in math.	−.194	.298	−5.78	
Total variance accounted for			65.80	

that less progress has been made through the curriculum by the age of thirteen. A late age of leaving might also account for it on the ground that there is still a lot of schooling to come after the age of thirteen.

Summary

The number of subjects studied by pre-university students studying mathematics ranges from an average of three in England to nine or more in several other systems of education. When a comparison is made between the mean scores of mathematics students from those systems where eight or more subjects are studied and those where four or fewer are studied, there is no significant difference in score. The conclusion 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, must be of interest to educational policy makers and planners in those countries where on average only few subjects are studied. In those countries where more subjects are studied, the age of terminating secondary schooling tends to be higher, and those countries where the age of terminating secondary schooling is lower tend to be those where the mandatory age of entry to school is lower.

The mandatory age of entry to school is five in England and Scotland, seven in Sweden and Finland, and six in the other systems paticipating in this study. The different degrees of pre-school attendance in the different systems are pointed out. When a comparison of mean scores of 13-year-old students with different ages of entry is made, differences are in favour of those entering at the age of six, but it must be remembered that the six year of entry scores are very heterogeneous. The average of the 13-year-old scores in Sweden and Finland (the latter, unweighted scores) is considerably lower than the average of the 13-year-olds with an age of entry of either six or five years.

Again, although it is easy to pick out pairs of countries to demonstrate that earlier age of entry would mean higher scores, the overall conclusion must be that age of entry at five or six is not associated with mathematics score at age 13. The extra year of schooling employed by those entering at five would not appear to be of consequence as far as progress in mathematics is concerned, whereas

the loss of a year's schooling between six and seven appears to have a detrimental effect.

Although it would appear that children from professional and white-collar social groups benefit more from early entry to school than do children from farmer and blue-collar social group, it is difficult to draw firm conclusions because of the heterogeneity of scores within each of the age of entry groups. However, this finding is not surprising, since it is to be expected that higher social group parents are likely to take more advantage than lower social group parents in a system with a fixed age of entry, since they are geared to that age of entry. It would be interesting to examine whether lower social group children really did score higher when given the chance to have earlier entry to school than some of their peers within a country.

It is, however, clear that to make the mandatory age of entry to school earlier will not, in itself, improve performance. It is what happens in that extra initial year which is important and it is the qualitative differences which must now be the subject of more systematic research.

In an attempt to discover if other aspects of school organization were likely to be of more importance when trying to account for differences between countries in scores of 13-year-olds, certain features were selected where there was known to exist considerable difference in practice between countries. The features chosen were number of subjects studied in the grade where most 13-year-olds were to be found in the school system, pre-service training of teachers, hours school per week and student's opportunity to learn the items on the test (i.e. the student's programme). Two other variables which pertain to some extent to the school and to some extent to the home were also chosen. They were "interest in mathematics" and "hours homework per week".

The correlations between these variables and national mean scores provided evidence of association. The regression equations suggest that the strongest evidence of association lay between the mean scores and the amount of pre-service teacher training, the amount of homework and the extent of the opportunity to learn. Evidence of association is not of itself evidence of a causal relation but it seems reasonable enough to think that in these cases the relation is causal.

From other national research (cf. Peaker, 1967) it is known that for primary school children within England school variables account

for about only twenty percent of the variance, whereas home variables (including parental attitudes and aspirations as well as socio-economic variables) account for about fifty percent of the variance. It is therefore suggested that in future international research school variables should be taken in conjunction with home variables when trying to account for differences between countries. It may turn out, of course, for home variables that unlike their contribution within countries, their contribution between countries is small.

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.

References

Anderson, C. A. *Methodology of Comparative Education,* International Review of Education, Volume VII, No. 1, 1961.

Anderson, Irving H. *Comparisons of the Reading and Spelling Achievement and Quality of Handwriting of Groups of English, Scottish and American Children.* Co-operative Research Project, No. 1903, University of Michigan, 1964.

Annuaire statistique de l'enseignement, Tome V. Belgique, 1960–61.

Bereday, G. F. Z. *Comparative Method in Education,* Holt, Rinehart and Winston, 1964.

Blandford, J. S. *Standardized Tests in Junior Schools with Special Reference to the Effects of Streaming on the Constancy of Results.* In: Brit. J. Ed. Psych. Volume XXVIII, 1958, pp. 170–173.

Boalt, G. and Husén, T. *Skolans Sociologi,* Almqvist and Wiksell, 1964.

Carroll, J. B. *Research on Teaching Foreign Languages.* In: Handbook of Research on Teaching, edited by N. L. Gage. Rand McNally & Co., Chicago, 1963.

Carnegie Quarterly. *The Gross Educational Product: How Much Are Students Learning?* Volume XIV, No. 2, 1966.

Dahllöf, U. *Kraven på Gymnasiet,* Stockholm, 1963.

Dahllöf, U., Zetterlund, S. and Öberg. H. *Secondary Education in Sweden.* Almqvist and Wiksell, Uppsala, 1966.

Dancy, J. *The Public Schools and the Future,* Faber and Faber, London, 1963.

Daniels, J. C. *Some Effects of Sex Segregation and Streaming on the Intellectual and Scholastic Development of Junior School Children.* Unpub. Ph.D. thesis, Nottingham University, 1959.

Daniels, J. C. *The Effects of Streaming in the Primary School: 2. A Comparison of Streamed and Unstreamed Schools.* In: Brit. J. Ed. Psych. Volume XXXI, pp. 119–127, 1961.

Douglas, J. W. B. *The Home and the School.* McGibbon and Kee, London, 1964.

Ekström, R. B. *Experimental Studies of Homogeneous Grouping: a Critical Review.* In: Sch. Rev. LXIX, pp. 216–226, 1961.

Emmett, W. G. *An Inquiry into the Prediction of Grammar School Success,* University of London Press, London, 1945.

Foshay, A. W. (ed.) *Educational Achievements of 13-year-olds in Twelve Countries,* Unesco Institute for Education, Hamburg, 1962.

Gatfield, R. A. *An Experimental Investigation into Certain Aspects of Streaming in a Primary School,* University of Southampton, 1958.

Goldberg, M. L., Passow, A. H. and Justman, J. *The Effects of Ability Grouping,* Teachers College Press, New York, 1966.

Goodlad, J. I. and Anderson, R. H. *The Non-Graded Elementary School.* Harcourt, Brace and World, Inc., New York, 1963.

Halsey, A. H. *Ability and Educational Opportunity, O.E.C.D.,* 1961.

Harbison, F. and Myers, C. A. *Education, Manpower and Economic Growth: Strategies of Human Resource Development.* McGraw Hill, New York, 1964.

Hitpass, J. *Bericht über eine 6-jährige Bewerbungskontrolle von Aufnahmeprüfungen und Testprüfungen.* In: Psychologie, Volume X, 1960, pp. 211–218.

Hotyat, F. *Les Examens : Les Moyens d'Evaluation dans L'Enseignement,* Bourrelier, 1962.

Husén, T. *Tonårsskola i utbildningssamhälle.* Almqvist and Wiksell, Stockholm, 1962.

Husén, T. *Problems of Differentiation in Swedish Compulsory Schooling,* Svenska Bokförlaget, Stockholm, 1962 *a.*

Husén, T. A working paper for an expert meeting at the International Institute for Educational Planning on Quality in Education. Paris, 1966.

Husén, T. *The Relation between Selectivity and Social Class in Secondary Education.* In: International Journal for the Educational Sciences. Volume I, pp. 17–27. Pergamon Press, Ltd., 1966.

Husén, T. *et al. International Project for the Evaluation of Educational Achievement. Phase I: International Study of Achievement in Mathematics: A Comparison of Twelve Countries.* Volumes I and II. Almqvist and Wiksell, Stockholm, and John Wiley & Sons, Inc., New York, 1967.

Husén, T. and Svensson, N.-E. *Pedagogic Milieu and Development of Intellectual Skills.* In: School Review, Chicago, 1960.

Jackson, B. *Specialization–how much and how soon?* In: *Where,* Advisory Centre for Education, London, March, 1966.

Khan, M. *A Study of Emotional and Environmental Factors Associated with Backwardness.* Unpublished Ph.D. thesis. University of London, 1954.

King, E. J. *Educational Progress and Social Problems in Japan.* In: Comparative Education, Volume I, No. 2, 1965.

Kish, L. *Survey Sampling.* John Wiley & Sons, Inc., New York, 1965.

Marklund, S. *Scholastic Attainments as Related to Size and Homogeneity of Classes.* Kungl. Skolöverstyrelsen, Stockholm, 1962.

Mogstad, Lars P. *Et tillskud til klarlegging av sentraliseringsspørsmålet i landsfolkeskulen.* In: Forskning og danning, 1958.

Passow, A. H. *The Maze of the Research on Ability Grouping.* In: Educational Forum, Volume XXVI, March 1962, pp. 281–288.

Peaker, G. F. *Appendix IV in Volume 2 of Children and their Primary Schools.* A Report of the Central Advisory Council for Education (England) H.M.S.O., January, 1967.

Pedley, R. *The Comprehensive School.* Pelican, London, 1963.

Pidgeon, D. A. *A Comparative Study of Basic Attainments.* In: Educational Research, Volume I, No. 1, 1958.

Pidgeon, D. A. *School Type Differences in Ability and Attainment.* In: Educational Research, Volume I, No. 3, 1959.

Pidgeon, D. A. *A Comparative Study of the Dispersion of Test Scores.* In: Educational Achievements of 13-year-olds in Twelve Countries, Unesco Institute for Education, Hamburg, 1962.

Pidgeon, D. A. *Date of Birth and Scholastic Performance.* In: Educational Research, Volume VIII, No. 1, 1965.

Pidgeon, D. A. *Education and the Concept of Intelligence.* Paper read at the Conference of the Association of Educational Psychologists, London, April, 1966.

Pidgeon, D. A. (ed.) *Achievement in Mathematics. A National Study in Secondary Schools.* N.F.E.R., 1967.

Postlethwaite, T. N. *Organisational Differentiation and School Achievement: A Preliminary Study of Mathematics Performance in Different School Systems.* Unpub. Fil. Lic. thesis, Stockholm University, 1965.

Rasmussen, M. and Prete, L. (eds.) *Toward Effective Grouping.* Association for Childhood Education International, Washington, D.C., 1962.

Robbins Report *Higher Education* Appendix I, H.M.S.O., 1963.

Rudd, W. G. A. *The Psychological Effects of Streaming by Attainment.* In: Brit. J. Ed. Psych., Volume XXVII, 1958, pp. 47–60.

Svensson, N.-E. *Ability Grouping and Scholastic Achievement.* Almqvist and Wiksell, Stockholm, 1962.

Undeutsch, U. *Auslese für und durch die höhere Schule.* In: Bericht über den zweiundzwanzigsten Kongress der deutschen Gesellschaft für Psychologie, Göttingen, 1960.

Wall, W. D., Olson, W. C. and Schonell, F. C. *Failure in School.* Unesco Institute for Education, Hamburg, 1962.

Wolf, R. M. *Data Bank Manual.* International Study of Achievement in Mathematics: A Comparison of Twelve Countries. Almqvist and Wiksell, Stockholm, 1967.

Yates, A. and Pidgeon, D. A. *Admission to Grammar Schools,* Newnes, London, 1957.

Yates, A. *Grouping in Education.* Almqvist and Wiksell, Stockholm, and John Wiley and Sons, Inc., New York, 1966.

Appendix

Table A.1. *Participants in I.E.A.*

Representing National Centers

Australia

Dr. J. P. Keeves, Australian Council for Educational Research, Hawthorn E2 Victoria.

Belgium

Monsieur F. Hotyat, Institut Supérieur de Pédagogie de Morlanwelz, Morlanwelz, Hainaut.

England

Dr. W. D. Wall, Mr. D. A. Pidgeon, Dr. K. M. Miller, The National Foundation for Educational Research in England and Wales, 79, Wimpole Street, London, W.1.

Federal

Republic of Germany Professor Dr. W. Schultze, Deutsches Institut für Internationale Pädagogische Forschung, Frankfurt/M. Schlossstrasse 29.

Finland

Professor M. Takala, Mr. O. Ylinentalo, Center for Educational Research, University of Jyväskylä, Jyväskylä.

France

Professor G. Mialaret, Université de Caen, Faculté des Lettres et Sciences Humaines, Caen.

Israel

Dr. M. Smilansky, Henrietta Szold Institute for Research in the Behavioural Sciences, 80 Haneviim Street, Jerusalem.

Japan

Professor S. Sakakibara, Dr. T. Harada, Dr. M. Kojima, Dr. S. Kubo, National Institute for Educational Research, 6-5-22 Shimomeguro, Meguro-ku, Tokyo.

Netherlands

Professor S. Wiegersma, Dr. M. Groen, Nederlands Institut voor Praeventieve Geneeskunde, 56, Wassenaarseweg, Leiden.

Scotland

Dr. D. A. Walker, The Scottish Council for Research in Education, 46, Moray Place, Edinburgh 3.

Sweden

Professor Dr. T. Husén, Dr. S. Henrysson, Mr. L.-M. Björkquist, School of Education, University of Stockholm, Stockholm 34.

U.S.A.

Professor C. A. Anderson, Professor B. S. Bloom, Department of Education, University of Chicago, 5835, Kimbark Avenue, Chicago 37, Illinois. Professor A. W. Foshay, Professor R. L. Thorndike, Teachers College, Columbia University, New York.

Consultants

Test Editors

Professor R. L. Thorndike, Mr. D. A. Pidgeon.

International Sampling Referee

Mr. G. F. Peaker, C. B. E., Grasmere, Ambleside, Westmorland, England.

Comparative Educationist

Mr. R. F. Goodings, University of Durham Department of Education, Old Elvet, Durham.

Data Processor

Dr. R. M. Wolf, University of Chicago.

Coordinator

Mr. T. N. Postlethwaite placed at the Unesco Institute for Education, Hamburg.

Table A.2. *Summary of topics for populations 1 a and 1 b.*

Subject matter	Objectives	Importance
000 ARITHMETIC		
001 Reasonable competence in the 4 operations on natural numbers	A, B	3
002 Ability to carry out simple operations involving decimal fractions	A, B	3
003 Ability to carry out simple operations involving simple vulgar fractions	A, B	2
004 Understanding the concept of fractions (vulgar and decimal)	C, D	3
005 Application of (001)-(004) to everyday life situations	C, D	3
006 Measurement of quantities, including length, area, volume capacity, time, speed and money	A, B	3
007 Notion of ratio and proportion, including percentages	A, C	3
008 Notion of arithmetical mean	A, C	2
009 Interpreting and making of simple practical graphs and tables	A, B, C	3
010 Intuitive understanding of properties of operations, i.e. associative, distributive, commutative laws	A, D	2
011 Expression of these laws by means of letters	B, C	1
012 Prime factors, divisors and multiples	A, B	2
013 Notions of powers and simple calculations of area and volume	A, C	2
014 Notions of number systems other than the decimal system	A, D, E	2
015 Notions of square roots	A	1
100 ALGEBRA		
101 Notions of positive and negative numbers/graphical representation	A, C	3
102 Extension to all positive and negative rational numbers of the four fundamental operations	A, B	2
103 Negative and zero exponents	A, C	1
104 Formulae and algebraic expressions	A, C	3

Table A.2. (*Continued.*)

Subject matter	Objectives	Importancee
105 Numerical evaluation of these formulae and algebraic expressions	B	3
106 Operations with polynomials and monomials	A, B	2
107 $(x+y)^2$, $(x-y)^2$ $(x+y)(x-y)$	A, B	1
108 Notions of equation	A	3
109 Equations of the first degree with numerical coefficients	B	3
110 Simple problems using (109)	C	3
111 Simple systems of linear equations with two unknowns	A, B	1
112 General (modern) notions of functions	A, C	2
113 Graphical representation of the functions of the type: $y=ax$; $y=ax+b$; $y=a/x$, $y=ax^2$	B, C	2
114 Elementary notions of sets	A, C	3
200 GEOMETRY		
201 Intuitive treatment of some geometrical figures: angle, triangle, square, parallelogram, rhombus, trapezium, circle	A	3
202 Intuitive treatment of: straight line, opposite angles, perpendicular and parallel	A	3
203 Intuitive treatment of symmetry and congruence	A	3
204 Intuitive treatment of translation and rotation	A	1
205 Measurement of distance and angles	A, B	3
206 Simple constructions with graduated ruler, straight edge, compasses, protractor, etc.	B	3
207 Intuitive treatment of similarity. Scale drawing	A, B, C	3
208 Properties of simple solids	A	2
209 Calculation of area and volume	B, C	2
210 Simple deductive reasoning based on the following: (a) properties of angles determined by 2 parallel lines, cut by a transversal and the sum of the angles in a triangle; (b) symmetry of isosceles triangle and rhombus; (c) fundamental conditions of congruence of 2 triangles (SSS, SAS); (d) inequality in triangles; (e) characteristic properties of the parallelogram.	A, D, E	3
211 Simple deductive reasoning based on the following: Properties of the inscribed angle of a circle	A, D, E	1
212 The theorem of Pythagoras for solving simple practical problems	A, B, C	2

Table A.3. *Summary of topics for population 3.*

Subject matter		Objectives	Importance
1.0	SETS, RELATIONS AND FUNCTIONS		
1.1	Sets		
	Notion of Sets		
	Intersection of Sets		
	Union of Sets	A, C	3
	Inclusion of Sets		
1.2	Relations and Functions		
	Condition in 2 variables		
	Sets of ordered pairs, relations	A, C, D	3
	Functional relations, etc.		
2.0	ARITHMETIC		
2.1	General treatment of number systems in terms of letters	A, B, C	3
2.2	Natural numbers	A, D	1
2.3	Integers	A, D	1
2.4	Real numbers	A, B	2
2.5	Complex numbers	A, B, D	1
3.0	ALGEBRA		
3.1	Polynomials	A, B	2
	Operations and Factorization		
3.2	Equations and Inequalities	A, B, C, D	3
3.3	Irrational equations	A, B, C	1
3.4	Systems of equations	A, B, C, D	3
3.5	Matrices and determinants	A, B	1
4.0	ELEMENTS OF ANALYSIS		
4.1	Polynomial functions	A, B2	3
4.2	Rational functions	A, B	2
4.3	Irrational functions	A, B	1
4.4	Circular functions	A, B, C, D	3
4.5	Inverse-circular functions	A, B	1
4.6	Logarithmic and exponential functions	A, B, C, D	3
4.7	Limits	A, B, C, D	3
4.8	Continuity	A, D	2
4.9	Derivatives	A, B, D	3
4.10	Integrals	A, B, D	2
4.11	Series	A, B	1
4.12	Differential equations	A, B, C	1
5.0	GEOMETRY		
5.1	Geometry mainly according to Euclid	A, B, D, E	2
5.2	19th-century geometry (projective, affine, etc.)	A, B, D	1

Table A.3. (*Continued.*)

Subject matter		Objectives	Importance
5.3	Trigonometry	A, B, C	1
5.4	Analytical geometry	A, B, C, D	2
5.5	Vectors	A, B, C	1
6.0	PROBABILITY AND STATISTICS		
6.1	Descriptive statistics	A, B, C	2
5.2	Probability	A, B, C, D	1
6.3	Distribution	A, D	1
6.4	Statistical inference	A	1
7.0	LOGIC		
7.1	Elementary formal logic	A, C, D	1
7.2	Deductive systems	A, D	1
8.0	HISTORY OF MATHEMATICS	A	1
9.0	ADDITIONAL TOPICS		

Table A.4. *Regression scaling of 1b and 3a onto the 3b scale.*

	Regression of test 3 on 3b score $y=a+b_x$		Regression of test 5 3b score $y=a+b_x$		1b Mean test 3	3a Mean test 5	3b Mean total	1b Mean scaled	3a Mean scaled
Belgium	−0.011	2.041	7.215	2.072	4.964	14.905	24.30	10.12	38.10
England	−2.109	2.095	8.398	2.128	4.156	14.659	22.10	6.60	39.59
Finland	−0.940	2.027	6.825	2.034	4.308	12.942	22.50	7.79	33.15
Germany	3.083	1.928	10.340	1.875	4.704	12.381	27.65	12.15	33.55
Japan	−6.268	2.491	8.046	2.234	6.104	14.326	25.36	8.94	40.05
Scotland	−2.029	2.053	7.926	2.160	3.499	12.424	20.77	5.15	34.76
Sweden	−0.488	1.553	9.028	1.758	2.151	12.436	12.69	2.85	30.89
U.S.A.	−1.105	1.848	4.375	2.146	2.244	7.066	8.63	3.04	19.56

Table A.5. *Indices of social bias.*

Country	Pop. 3a	3b	Country	Pop. 3a	3b	Country	Pop. 3a	3b
Australia	4.7	—	Finland	6.0	3.7	Netherlands	12.3	—
Belgium	3.6	7.3	France	17.3	—	Scotland	10.4	5.7
England	16.2	24.5	Israel	3.6	—	Sweden	2.1	7.0
Fed. Rep. of Germany	45.3	56.4	Japan	6.0	2.9	U.S.A.	1.9	1.0

Table A.6. *Means, standard deviations and N's for total mathematics score, lower mental process and higher mental process by sub-sample.*

(Population 1b)

Country	Total math. score			Lower mental process score			Higher mental process score		
	M	S.D.	N	M	S.D.	N	M	S.D.	N
AUSTRALIA									
Subsample 1	17.58	12.34	770	13.83	9.53	770	3.76	3.65	770
Subsample 2	18.32	11.30	769	14.86	9.03	769	3.46	3.28	769
Subsample 3	21.44	12.42	770	16.90	9.78	770	3.59	3.71	769
Subsample 4	18.16	13.04	770	14.56	10.10	770	3.59	3.71	770
Average	18.88	12.28	770	15.04	9.61	770	3.84	3.53	770
BELGIUM									
Subsample 1	26.43	12.96	661	12.15	10.09	661	5.28	3.72	661
Subsample 2	32.37	14.05	661	17.92	11.33	661	5.06	4.30	661
Subsample 3	34.04	12.74	661	26.95	9.55	661	7.13	3.99	661
Subsample 4	28.90	15.25	661	22.53	11.51	661	6.38	4.45	661
Average	30.43	13.75	661	23.98	10.41	661	6.46	4.11	661
ENGLAND									
Subsample 1	25.25	18.45	793	19.22	14.21	793	6.03	4.86	793
Subsample 2	22.28	18.81	789	16.95	14.33	789	5.28	4.93	789
Subsample 3	23.69	177.3	773	18.31	13.44	773	7.38	4.86	773
Subsample 4	23.71	19.03	793	18.23	14.73	793	5.47	4.91	793
Average	23.76	18.53	787	18.20	14.19	787	5.56	4.90	787
FINLAND									
Subsample 1	25.98	9.59	210	19.37	6.88	210	6.07	3.55	210
Subsample 2	25.51	8.56	210	19.48	6.61	210	6.03	2.89	210
Subsample 3	26.43	9.73	210	20.19	7.40	210	6.25	3.14	210
Subsample 4	27.55	10.39	210	20.25	7.69	210	7.30	3.43	210
Average	26.37	9.57	210	19.82	7.15	210	6.54	3.25	210
FRANCE									
Subsample 1	19.10	13.95	922	15.18	10.54	922	3.91	3.95	922
Subsample 2	22.87	12.99	924	18.12	9.77	924	4.76	3.96	924
Subsample 3	21.09	12.92	710	16.70	9.68	710	4.39	3.88	710
Subsample 4	20.76	13.05	893	16.75	9.92	893	4.01	3.76	893
Average	20.96	13.23	862	16.69	9.98	862	4.27	3.89	862
GERMANY									
Subsample 1	23.95	11.67	1119	18.72	8.94	1119	5.24	3.43	1119
Subsample 2	23.22	12.76	1119	17.64	9.52	1119	5.58	3.78	1119
Subsample 3	27.80	10.98	1119	21.88	8.13	1119	5.92	3.73	1119
Subsample 4	26.85	11.41	1119	20.69	8.59	1119	6.16	3.43	1119
Average	25.45	11.70	1119	19.73	8.80	1119	5.72	3.59	1119

Table A.6. (*Continued.*)

Country	Total math score			Lower mental process score			Higher mental process score		
	M	S.D.	N	M	S.D.	N	M	S.D.	N
HOLLAND									
Subsample 1	21.89	12.11	361	16.88	9.03	361	5.01	3.74	361
Subsample 2	19.53	12.71	361	15.01	9.70	361	4.52	3.61	361
Subsample 3	21.79	10.68	361	16.72	8.04	361	5.03	3.46	361
Subsample 4	22.52	12.94	361	17.40	9.62	361	5.13	3.80	361
Average	21.43	12.12	361	16.51	9.10	361	4.92	3.68	361
ISRAEL									
Subsample 1	34.95	14.20	834	26.31	10.16	834	8.64	4.75	834
Subsample 2	31.11	13.88	759	23.86	9.84	759	7.35	4.65	759
Subsample 3	29.76	16.60	805	22.83	12.10	805	6.93	5.08	805
Subsample 4	33.35	13.99	834	25.38	10.10	834	7.97	4.62	834
Average	32.29	14.67	3232	24.59	10.55	3232	7.70	4.77	3232
JAPAN									
Subsample 1	32.38	17.00	512	25.52	12.56	512	6.86	5.07	512
Subsample 2	31.28	16.92	513	24.40	12.54	513	6.87	5.02	513
Subsample 3	31.11	16.73	512	24.54	12.33	512	6.57	4.98	512
Subsample 4	29.87	16.94	512	23.69	12.51	512	6.19	5.07	512
Average	31.16	16.90	2050	24.54	12.48	2050	6.62	5.03	2050
SCOTLAND									
Subsample 1	23.72	15.22	1443	18.48	11.75	1443	5.24	4.08	1443
Subsample 2	22.72	16.60	1440	17.62	12.70	1440	5.10	4.48	1440
Subsample 3	20.45	15.92	1440	15.96	12.39	1440	4.49	4.17	1440
Subsample 4	22.32	15.03	1395	17.57	11.91	1395	4.75	3.83	1365
Average	22.31	15.69	1425	17.41	12.18	1425	4.90	4.14	1425
SWEDEN									
Subsample 1	15.97	10.81	727	11.55	8.32	727	3.42	3.38	727
Subsample 2	14.32	10.35	656	11.32	8.12	656	3.00	3.05	656
Subsample 3	15.56	11.48	737	12.14	8.89	737	3.41	3.36	737
Subsample 4	15.20	10.73	708	12.07	8.30	708	3.12	3.27	708
Average	15.26	10.83	707	12.02	8.41	707	3.24	3.26	707
UNITED STATES									
Subsample 1	17.42	13.49	1622	14.20	10.42	1622	3.22	3.80	1622
Subsample 2	19.14	12.90	1639	15.35	9.93	1639	3.79	3.73	1639
Subsample 3	18.23	12.98	1662	14.69	10.09	1662	3.54	3.64	1662
Subsample 4	16.61	13.89	1621	13.53	10.66	1621	3.07	3.92	1621
Average	17.85	13.32	6544	14.44	10.28	6544	3.40	3.77	6544

Table A.7. *Means, standard deviations and N's for total mathematics score, lower mental process and higher mental process by sub-sample.*

(Population 1a)

Country	Total math. score			Lower mental process score			Higher mental process score		
	M	S.D.	N	M	S.D.	N	M	S.D.	N
AUSTRALIA									
Subsample 1	20.37	14.60	729	16.07	11.34	729	4.30	3.97	729
Subsample 2	19.33	13.48	729	15.56	10.59	729	3.78	3.69	729
Subsample 3	22.21	14.31	729	17.59	11.05	729	4.62	4.03	729
Subsample 4	18.82	13.65	729	15.04	10.65	729	3.77	3.76	729
Average	20.18	14.01	729	16.06	10.09	729	4.12	3.86	729
BELGIUM									
Subsample 1	23.67	14.11	387	19.04	11.02	387	4.63	3.92	387
Subsample 2	28.53	14.82	433	22.34	11.16	433	6.19	4.34	433
Subsample 3	30.94	15.15	433	24.41	11.70	433	6.52	4.16	433
Subsample 4	27.82	15.99	433	21.58	12.05	433	6.25	4.62	433
Average	27.74	15.02	422	21.84	11.48	422	5.90	4.26	422
ENGLAND									
Subsample 1	33.02	15.89	736	25.00	12.14	736	8.02	4.55	736
Subsample 2	13.88	12.47	776	10.51	9.13	776	3.78	3.80	776
Subsample 3	15.10	11.23	750	11.03	8.77	750	4.07	3.09	750
Subsample 4	19.19	17.53	750	14.77	13.61	750	4.43	4.53	750
Average	19.31	16.97	753	14.79	13.17	753	4.53	4.42	753
FINLAND									
Subsample 1	24.55	10.07	187	18.21	7.17	187	6.34	3.75	187
Subsample 2	23.24	9.16	187	17.85	7.17	187	5.39	2.91	187
Subsample 3	24.30	9.67	187	18.40	7.39	187	5.90	2.96	187
Subsample 4	24.16	10.52	187	17.89	7.74	187	6.28	3.43	187
Average	24.06	9.85	187	18.09	7.36	187	5.98	3.86	187
FRANCE									
Subsample 1	13.96	10.40	589	11.31	8.28	589	2.64	2.84	589
Subsample 2	22.02	13.60	662	17.39	10.30	662	4.62	4.05	662
Subsample 3	19.69	13.18	523	15.47	9.83	523	4.22	3.90	523
Subsample 4	17.61	12.30	636	14.32	9.41	636	3.29	3.57	636
Average	18.32	12.37	602	14.62	9.45	602	3.70	3.59	602
HOLLAND									
Subsample 1	24.59	15.62	107	18.95	11.70	107	5.64	4.35	107
Subsample 2	24.18	18.36	107	18.53	13.73	107	5.65	5.01	107
Subsample 3	22.72	14.22	107	17.48	10.82	107	5.24	3.90	107
Subsample 4	23.98	15.46	107	18.45	11.65	107	5.53	4.43	107
Average	23.86	15.91	107	18.35	11.98	107	5.52	4.42	107

Table A.7. (*Continued.*)

Country	Total math score			Lower mental process score			Higher mental process score		
	M	S.D.	N	M	S.D.	N	M	S.D.	N
JAPAN									
Subsample 1	32.38	17.00	512	25.52	12.56	512	6.86	5.07	512
Subsample 2	31.28	16.92	513	24.04	12.54	513	6.87	5.02	513
Subsample 3	31.11	16.73	512	24.54	12.33	512	6.57	4.98	512
Subsample 4	29.87	16.94	512	23.68	12.51	512	6.19	5.07	512
Average	31.16	16.90	512	24.54	12.48	512	6.62	5.03	512
SCOTLAND									
Subsample 1	20.58	14.80	1326	15.99	11.40	1326	4.59	3.99	1326
Subsample 2	20.17	15.88	1323	15.73	12.35	1323	4.44	4.18	1323
Subsample 3	18.03	14.90	1323	14.11	11.69	1323	3.93	3.89	1323
Subsample 4	17.43	13.00	1284	13.72	10.40	1284	3.71	3.37	1284
Average	10.05	14.64	1314	14.88	11.46	1314	4.17	3.86	1314
SWEDEN									
Subsample 1	16.30	10.94	658	12.81	8.39	658	3.49	3.43	658
Subsample 2	15.21	10.07	595	12.08	7.93	595	3.13	2.99	595
Subsample 3	15.63	11.64	663	12.18	9.03	663	3.44	3.39	663
Subsample 4	15.66	10.60	637	12.49	8.22	637	3.17	3.23	637
Average	15.70	10.81	638	12.39	8.39	638	3.31	3.26	638
UNITED STATES									
Subsample 1	16.81	12.73	1540	4.33	6.14	1540	1.83	4.02	1540
Subsample 2	18.00	12.85	1566	5.48	10.38	1566	5.30	3.76	1566
Subsample 3	17.00	12.63	1582	8.20	5.19	1582	2.93	3.43	1582
Subsample 4	15.62	12.49	1543	5.11	10.47	1543	.25	3.78	1543
Average	16.15	13.34	1558	13.06	10.36	1558	3.09	3.72	1558