

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			Age s.d.
	M	s.d.	N	
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.

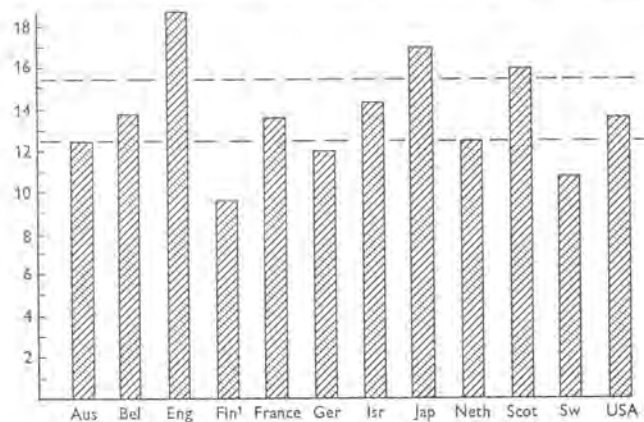
Within a year group setting (the grouping of students for specific subjects or activities only according to their ability or achievement) will have similar effects to streaming in increasing the spread of scores of the age group on any achievement criterion.

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 16 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 stand-

ard 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 empirically 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.

Country	(Population 1a)			Age s.d.	Age s.d. (1b)
	Total mathematics score				
	M	s.d.	N		
Australia	20.18	14.01	2916	3.5	7.7
Belgium	27.74	15.02	1686	3.3	8.3
England	19.31	16.97	3012	3.3	4.2
Finland [#]	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

[#] 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 (1a) standard deviation will be more strongly associated with the index of grade repeating that has been chosen than the standard deviation of the grade popu-

lation (1b). The correlation is $-.05$, which although negative is less so than the correlation of $-.53$ in Population 1b 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 3
 - (Indicate in which groups under "Comment")
 - 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		Number of schools	Population 1b Ability grouping		Number of schools
	(1)	(2)		(1)	(2)	
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 1b and 1a 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 weights³ 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

* 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 sys-

tem 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.