Chapter 8

Change and Stability in RPM Scores Over Culture and Time: The Story at the Turn of the Century

John Raven

Introduction

This chapter is based upon a paper published in *Cognitive Psychology* (Raven, 2000b), which was itself based on material previously published in the *Manuals* for the *Raven Progressive Matrices* – see various reference entries for Raven, Raven, and Court. It offers a more extended summary of research relating to the stability and change in RPM scores over culture and time (and their causes) than was included in the *General Introduction* to this book.

Given that the tests have been in use for 70 years, distilling off the evidence bearing on the question of how similar are the norms for different cultural groups and how have they changed over time is not so easy as might be expected. The reasons why the data are not better than they are merit review because of their implications for future work in the area.

One reason why the task is so difficult is that, as Dahlstrom (1993) noted in an article appropriately titled "*Tests: Small samples, large consequences*" that most of the studies in the area are not only of poor quality but also conducted for other purposes. Thus most of the research in which the RPM have been used have sought to relate RPM scores to some other variable (such as educational or occupational performance) rather than to assemble basic normative data. And, both in these studies and in those which attempted to provide some kind of reference data, the researchers concerned have been relatively indifferent to the importance of sampling … even though this has major implications for the validity of the significance tests they sought to apply. Thus many researchers have tried to *explain* apparent cultural differences without first establishing just how large or pervasive those differences were.

Creeping Awareness of the Importance of Studying Change Over Time

However, so far as change over time is concerned, one reason why there is little adequate data is that, since most psychologists never even suspected that scores would increase over time, they not only did not think it was necessary to collect data which would bear on the question, they simply assumed that normative data collected in the past were still applicable. In short, they did not see any need to restandardise tests.

Another factor is, however, that the available evidence that scores were increasing over time was misinterpreted – as were the similar data available from numerous cross-sectional studies conducted with other tests – as evidence of declining ability after 20 years of age.

Although some researchers (e.g. Owens, 1966; Bouvier, 1969; Thorndike, 1975) did notice an apparent increase in scores on some components of "intelligence", the overall effect, when scores on these sub-tests were combined with the others, could hardly be described as "dramatic". Had the researchers concerned been in the habit of thinking in Spearman's terms – separating eductive and reproductive abilities in their minds, they might have noticed that it was essentially only one component of "intelligence" (namely, *eductive* ability) which was increasing and that the increase in such scores was indeed dramatic. (In this context we may note that, besides, in 1984 and 1987, documenting the rise more thoroughly than previous workers, Flynn's main contributions (eg Flynn, 2000) have in fact been to draw attention, first, to the *rate* of change in eductive ability, and, second, by forcefully raising the question of whether this increase has resulted in a genuine increase in *knowledge*, to underline the *differential* rate of change in eductive and reproductive abilities.)

Beyond these background constraints, there were also substantial methodological problems. First, to be meaningful, the data had to be sectioned by age, as in Table 8.1 (which will be explained more fully later). Second, the bimodal and skewed within-age distributions shown in Figure 8.1 (redrawn from J. Raven, 1981), combined with a scatter which varied with age (also illustrated in Figure 8.1), meant that the usual data reduction techniques (i.e. reduction of the data to means and standard deviations) could not meaningfully be adopted.

Nevertheless, despite the validity of most of these components in an explanation of psychologists' failure to notice the increase in scores that

Table 8.1. *Standard Progressive Matrices*
1979 British Percentile Norms for the Self-Administered or Group Test Among Young People (Smoothed)

	Age in years (months)																		
	6½	7	7½	8	8½	9	9½	10	10½	11	11½	12	12½	13	13½	14	14½	15	15½
	6(3)	6(9)	7(3)	7(9)	8(3)	8(9)	9(3)	9(9)	10(3)	10(9)	11(3)	11(9)	12(3)	12(9)	13(3)	13(9)	14(3)	14(9)	15(3)
	to	to	to	to	to	to	to	to	to	to	to	to	to	to	to	to	to	to	to
Percentile	6(8)	7(2)	7(8)	8(2)	8(8)	9(2)	9(8)	10(2)	10(8)	11(2)	11(8)	12(2)	12(8)	13(2)	13(8)	14(2)	14(8)	15(2)	15(8)
95	33	34	37	40	42	44	46	48	49	50	51	52	53	54	54	55	56	57	57
90	30	32	35	38	40	42	44	46	47	48	49	50	51	52	53	54	54	55	55
75	22	26	30	33	36	38	41	42	43	44	45	46	47	49	49	50	50	51	51
50	**16**	**19**	**22**	**25**	**31**	**33**	**36**	**38**	**39**	**40**	**41**	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**47**
25	13	14	15	17	22	25	28	32	33	34	36	37	38	39	41	42	42	42	42
10	10	12	12	14	16	17	19	23	27	29	31	31	32	33	35	36	36	36	36
5	9	10	11	12	13	14	15	17	22	24	25	26	27	28	29	30	33	33	33
n	112	138	148	174	153	166	198	172	194	187	164	164	174	185	180	196	189	191	171

Note. Based on a nationally representative sample of British schoolchildren, excluding those attending special schools (see Raven, 1981 for details). Younger and less able children were tested individually.

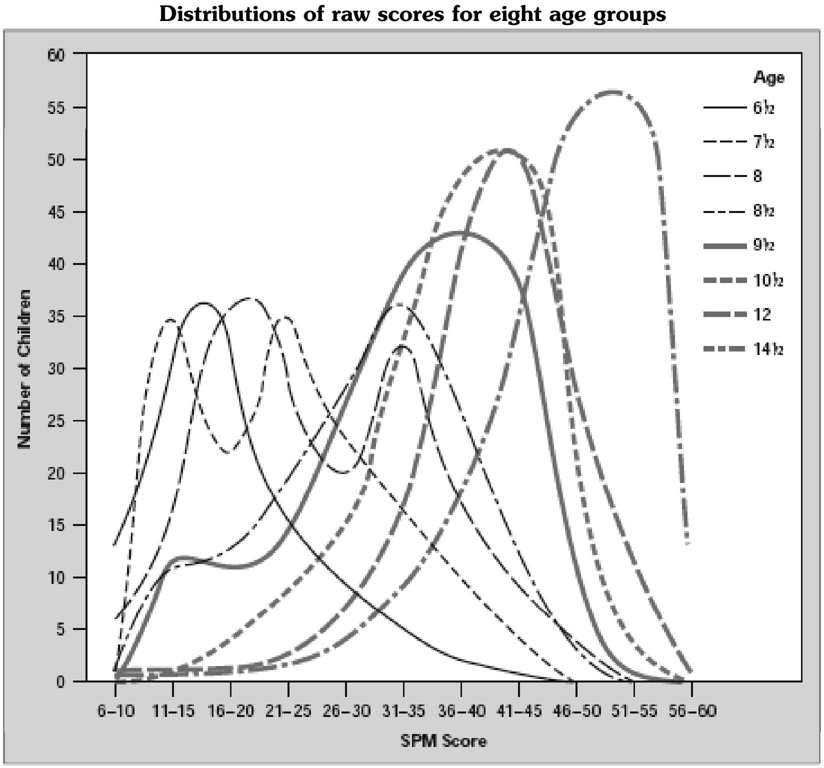
was occurring, the most important is probably that, until 1979, such data as were available would not really have led anyone to suspect an increase – even on *Progressive Matrices.* So let us now review such data as were available to shed light on changes over time and culture for young people and later do the same thing for the adult data.

Studies of Young People

1. Studies relating to change over time

The *Standard Progressive Matrices* (SPM) was first fully standardised by J.C. Raven on 1,407 children in Ipswich, England, in 1938 (J. C. Raven, 1941). The next substantial study (J. C. Raven & Walshaw, 1944) was conducted, not in order to produce norms for the RPM, but to gather equivalent data for the *Mill Hill Vocabulary Scale* (MHV). It was carried

Figure 8.1. *Standard Progressive Matrices*
1979 British Standardisation
Distributions of raw scores for eight age groups



out in a town not far from Ipswich, namely, Colchester, in 1943-44. The SPM norms obtained in that study were consistently two raw score points *lower* than the Ipswich norms. In 1952, Adams reported norms from 11,621 12-year-old children in Surrey, England. These data were, within the limits of sampling error, very similar to Raven's 1938 (Ipswich) norms. Tuddenham, Davis, Davison, and Schindler (1958), in one of the few studies which attempted to establish the appropriateness or otherwise of the British norms in the United States, tested several school classes of Californian children. They concluded that the British norms were acceptable. In 1963-65 Skanes tested 4,017 children aged 9 ½ to 14 years in St. John's, Newfoundland. The similarity between Skanes' results and the 1938 Ipswich norms is striking (J. Raven, 1981). Later, in 1967, in Corner Brook, Newfoundland Skanes tested the entire population (2,097) of children aged 10 ½ to 14 ½ years. The results consistently lagged *behind* the Ipswich norms. In 1972, Byrt and Gill (1973), working with the author, collected data from a nationally representative sample of 3,464 primary school children aged 5 ½ to 11 ½ years in the Republic of Ireland. The urban norms seemed to corresponded to the 1938 Ipswich norms, although the figures for rural areas lagged behind.

As late as 1979 – 40 years after the test was published – therefore, there was little to suggest a secular increase in scores. Quite the contrary: everything suggested stability.

From 1979 onwards the story began to change. In that year, Kratzmeier and Horn (1979) reported norms from a large German study which were well above those obtained in England in 1938. Melhorn's (1980) East German data were similar. The 1979 British norms, compiled (with the aid of a Social Science Research Council grant and assistance from the Government Office of Population Censuses and Surveys) from a carefully drawn sample designed to represent both the whole of Great Britain and the socioeconomic variance within it, appeared to be broadly similar to those obtained in the two German studies (J. Raven, 1981). Holmes (1980) reported results for British Columbia (Canada) which were similar to, if slightly lower than, the 1979 U.K. national norms. Both the Australian Council for Educational Research (see de Lemos, 1984, 1989) and the New Zealand Council for Educational Research (1984) reported closely corresponding results for their respective countries. Ferjencik (1985) reported data for the *Coloured Progressive Matrices* (CPM) for what was then Czechoslovakia which corresponded to a recently reported British study. Work carried out in the U.S. by J. Raven (2000a) revealed that, while the overall U.S. norms lagged behind these new international figures, the White norms did not. Zhang and Wang (1989) collected data for urban mainland China which showed that, despite what had been suggested by the high norms reported by Chan (1981, 1989) for Hong Kong, norms for a sample designed to be representative of urban mainland China corresponded closely to recent norms obtained elsewhere. More recently, similar data have been reported for Poland (Jaworowska & Szustrowa, 1991; J. Raven, J.C. Raven & Court, 1998c, 2000, updated 2004), Spain (J.C. Raven, Court, & J. Raven, 1995), further school districts of the U.S. (J. Raven & Court, 1989), Switzerland (Martinolli, 1990; Spicher, 1993) and India (Deshpande, in J. Raven , J.C. Raven, & Court, 2000).

Two observations may be interjected at this point. First, when reporting the results of the 1979 British standardisation, we ourselves (J. Raven, 1981), while noting the difference between the 1938 and 1979 norms, failed to comment on its *magnitude* and, overlooking the fact that the scores of the more able adolescents approached the maximum obtainable on the test, suggested that the increase had mainly occurred among the less able. Second, given the similarity in the norms reported by all the researchers listed in the last paragraph who published data from 1936 to 1979 and the similarity in the 1980s norms reported by the other authors whose work has been summarised, there was no hint that we might be looking at evidence of a *continuous* increase in scores over time. There could just have been a jump.

2. Geographical and cultural variance

The studies outlined thus far suggest that the norms for different populations are similar at a given point in time but had somehow jumped dramatically in the 1970s.

We will now summarise studies documenting variance in the norms for young people from different geographical areas and between cultures both as a topic in its own right and with a view to exploring what light they are able to shed on the changes over time. Studies revealing broad differences between countries will be reviewed first and followed by a review of studies of variance within countries.

As has been mentioned, Chan's Hong Kong norms exceed most of the norms already mentioned. However the norms which most significantly exceed them come from Taiwan (Miao & Huang, 1990; Miao, 1993). (To reduce the likelihood of an inappropriate interpretation being placed on these norms it should, be mentioned that, in the course of a visit to Taipai, it emerged that not only does the RPM play a major role in the rigid Taiwanese system of educational assessment and school placement, all teachers are supplied with copies of the test and encouraged to "train pupils in its use". It therefore seems likely that the "high" Taiwanese norms reveal, not the superior eductive ability of the Taiwanese, but rather, how small is the maximum effect of motivation and training.)

On the other hand, as also noted, norms for rural and isolated communities are typically lower than others. The previously mentioned norms for the Republic of Ireland and Newfoundland can, in this context, be seen to confirm this. Other low norms for what appear to be good samples of the relevant populations have been reported for Brazil (Angelini, Alves, Custodio, & Duarte, 1988), Turkey (Sahin & Duzen, 1994), Malaysia (Chiam, 1994, 1995), Puerto Rico (Kahn, Spears, & Rivera, 1977; J. Raven & Court, 1989), and a remote area in the mountains of Peru (see Munoz in Raven et al., 1998b).

As emphasised by J. Raven (1989), the "low" norms reported in most of these studies must be set in an appropriate context by observing that, with the notable exception of the Peruvian mountain norms, most are above the British 1938 norms. It follows that the factors that have been responsible for the shortly-to-be-discussed increases in scores over time could also have caused the differences between cultural groups.

More systematic studies of the variance between geographical, socioeconomic, and ethnic groups within countries were undertaken in the course of both the British and U.S. standardisations among young people. Because both the designs and the variables considered in these two studies were different they must be discussed separately.

The 1979 British standardisation

The 1979 British Standardisation was conducted in seven areas of the country which were chosen, under the guidance of the Government Office of Population Censuses and Surveys (OPCS), to represent all the types of area into which a cluster analysis of large amounts of demographic data had shown the socio-economic variance within the country could be classified (Webber, 1977). The types of area in which few people lived were over-sampled in order to have enough respondents to make it possible to break the data down by type of region. Later, the data were re-weighted to its correct proportions to give overall statistics. It was therefore possible to employ fairly sophisticated statistical procedures when analysing the data. Altogether 3,250 children aged 6 to 16 were tested.

This is a convenient point at which to explain the format in which the data will be displayed. Many authors present their data in terms of Deviation IQs with a mean of 100 and a standard deviation of 15. This process is, in general, unjustifiable for a number of reasons which include two that are important here: First, as was evident from Figure 8.1 the within-age score distributions for the RPM (and, according to a personal communication from Robert Thorndike, the subscales of the Stanford-Binet test) are generally not Gaussian and are, indeed, often bimodal. Second, it does not encourage enquiry into whether there may be differential trends at different ability levels.

To avoid these (and other) problems, the normative data for the RPM have always been presented in the form of tables showing the raw scores required to do better than 5%, 10%, 25%, 50%, 75%, 90% and 95% of the population of a particular age. Table 8.1, giving the overall results from the 1979 British norming study among young people, is typical of the output.

One further word of explanation is appropriate. In order minimise the – usually fairly large – effects of sampling error, the figures in these tables have typically been smoothed by drawing a graph for each percentile curve which, as far as possible, (a) equalises the deviations of the raw scores above and below the line, (b) gives most weight to the most reliable data (the data for the oldest and youngest age groups in a sample is typically distorted by factors such as advancement and retention policies in education), and (c) reflects the trends in the most reliable data.

The need to smooth the data to minimise the effects of random variance and especially sampling error may be seen from a glance at Figures 8.5 and 8.9. The populations on which these graphs are based are not correctly described as "samples" at all because they consist of virtually the entire populations of successive age cohorts of Belgian men. Nevertheless, failure to graph the data would obscure the general trends that are so striking and there would always be the temptation to focus on explaining fluctuations around the general trend which, given the huge numbers, would always turn out to be overwhelmingly "statistically significant". When the data are based on smaller numbers and sampling error comes into play, the results become even more irregular (illustrative data will be found in Table B1 in Raven (2000b).

The importance, of, when drawing these graphs, giving more weight to the most reliable data can be reinforced by noting that, if one has 50 people in an age category, the 5th and 95th percentiles are the scores lying half way between those obtained by the 2nd and 3rd person in the respective tails of the distribution. On their own, these estimates are therefore necessarily extremely unreliable.

The vitally important conclusions to be drawn from this digression are that the presentation of unsmoothed raw data can (i) lead to diversionary enquiries into the causes of chance fluctuations and (ii) when used as reference data against which to view the scores of individuals or experimental groups, to seriously misleading evaluations.

Table 8.1 (above) shows the (smoothed) raw scores corresponding to the 5^{th}, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th}, and 95^{th} percentiles for each age group in the 1979 British Standardisation.

When the 1979 U.K. data were broken down by a series of demographic variables (region, socio-economic status etc.) and analysed using multiple regression techniques (see Raven, 1981; 2000b) it emerged that 2.6 % of the SPM variance was accounted for by region. However, when the effect of Socio-Economic Status was partialled out, this dropped to 0.5%. Thus regional variation *per se* seems to be of little importance. SES on its own accounted for 4.8% of the variance. However, since age accounted for 46% of the variance, SES accounted for 8.9% of the variance which is not attributable to age. This is equivalent to a within-age correlation between Socio-Economic Status and the SPM of .30.

Population balance assessed via SES *is* therefore something that must be taken into account when comparing one set of results with another or when seeking to generalise from one population to another.

SPM score correlated .68 with age. Thus, more than half the variance was *not* "explained" by age. It is not, therefore, true that the tests simply measure "intellectual maturity."

As in the 1938 standardisation, Item Characteristic Curve (Item Response Theory or Rasch-type) based item analyses were carried out separately within each socioeconomic and age group. While the detailed figures take up too much space to present here, it may be noted (a) that the ICCs for individual items were remarkably similar to those published 40 years earlier and (b) that, as can be seen from the summary data presented in Table 8.2, the items scaled in much the same way for children from a variety of different backgrounds. . More recently, Vodegel-Matzen (1994) has shown that making the items more "realistic" (i.e. using hats, bananas etc. instead of abstract figures) while retaining the logic of the items makes the items easier for everyone, but changes neither the order of the items nor the order of individuals.

The conclusion is clear and vitally important: It is not possible to explain away differences in the mean scores of these groups on the grounds that, in any general sense, the test is "foreign to the way of thought of children from certain backgrounds." With certain important group and individual exceptions which will not be discussed here, the test generates orderly data which, on these grounds alone, must have some meaning. Differences between groups cannot be dismissed as "meaningless." They merit investigation and explanation.

U.S. Standardisations in the 1980s

Between 1983 and 1989 some 50 norming studies were carried out within school districts spread across the United States of America (J. Raven, 2000a). Within each district the sample was, as far as practicable, representative of the district. The specific sampling procedure employed varied from district to district, but, for reasons discussed in Appendix B, in no case were quota sampling procedures employed. (The sampling procedure adopted in each district is described in the previously mentioned publications.) Altogether more than 60,000 students aged 5 to 18 years were tested.

The norms which were obtained varied markedly from one school district to another and, within districts, between socioeconomic and ethnic groups.

As is illustrated in Tables 8.3 and 8.4 both ethnicity and socioeconomic status seemed to make independent contributions to the within-district variances. (It is important to note that the application of multiple regression

Table 8.2. *Standard Progressive Matrices*
1979 British Standardisation
Correlations Between Item Difficulties Calculated Separately for Young People from Different Socioeconomic Backgrounds
(Decimal points omitted and rounded to two decimal places)

SES	1 (High)	2	3	4	5	6	7	8 (Low)
1 (High)								
2	99							
3	99	99						
4	98	99	99					
5	97	98	99	99				
6	98	99	99	99	99			
7	95	96	98	98	99	98		
8 (Low)	95	96	98	98	99	99	99	

techniques is not strictly legitimate not only because the distributions are not Gaussian, but also because the independent variables are correlated with each other.) Accordingly, as shown in Table 8.4, the regressions were run twice with the independent variables entered in different orders: once with SES partialled out first and once with ethnicity partialled out first. Whichever way round the variables are entered, there is variance left to be explained by the other.

Differences between the norms for *school systems* catering for white students of differing socioeconomic status were as great as the ethnic differences within school districts.

The differences between the norms for school systems catering for different proportions of white, black and native American children seem to correspond to differences in published statistics on birth weight, infant mortality, and the incidence of serious childhood illness as published by the U.S. Bureau of the Census (United States Government, Bureau of the Census, 1984).

Within a number of school districts which had enough students of differing ethnicity to make the process legitimate, item analyses were run separately among different ethnic groups. One, fairly typical, example of the outcome is shown in Table 8.5. It follows from results like these (which duplicate those published by Jensen (1974) that the test works in the same way – measures the same thing – in each group. In addition, as illustrated in Figure 8.2, Hoffman (1983, 1990) demonstrated that the regression lines of RPM on various types of academic achievement for different ethnic groups were (to all intents and purposes) parallel – although having different intercepts. (Although the regression lines for mathematics shown in Figure 8.2 diverge while those for reading converge, these are only two examples. Overall, some diverge and some converge in such a way that it becomes clear that the general conclusion is that they are parallel.) Thus, while ethnic groups score at different levels on both achievement and matrix tests, the RPM has equal predictive validity within each group. Similar results were again reported by Jensen (1974).

Table 8.3. *Standard Progressive Matrices*
1986 Adolescent Percentile Norms for Ethnic Groups in Westown (U.S.) in the Context of 1979 British Data (Smoothed)

Percentile		Age in Years								
		12½	13	13½	14	14½	15	15½	16	16½
95	UK	53	54	54	55	56	57	57	–	–
	Anglo	51	52	53	54	55	56	57	58	59
	Asian	53	54	54	54	55	55	56	57	57
	Hisp	48	49	49	50	51	52	53	53	53
	Black	47	48	49	50	51	52	53	54	54
90	UK	51	52	53	54	54	55	55	–	–
	Anglo	50	51	52	53	54	55	56	57	57
	Asian	50	51	51	52	53	53	54	55	55
	Hisp	45	46	47	48	49	50	51	52	52
	Black	45	46	47	48	49	50	51	52	52
75	UK	47	49	49	50	50	51	51	–	–
	Anglo	46	47	47	48	50	52	53	54	54
	Asian	46	47	48	48	49	50	50	51	52
	Hisp	42	43	44	45	46	47	48	49	50
	Black	42	42	42	42	44	45	46	49	49
50	UK	42	43	44	45	46	47	47	–	–
	Anglo	41	42	43	44	47	48	48	48	49
	Asian	42	43	43	43	44	45	46	47	48
	Hisp	37	38	39	40	41	42	43	44	45
	Black	36	36	37	38	39	40	41	43	44
25	UK	38	39	41	42	42	42	42	–	–
	Anglo	37	38	39	40	42	44	45	45	45
	Asian	35	35	36	36	37	38	40	42	43
	Hisp	32	33	34	35	36	37	39	39	40
	Black	29	30	31	32	33	34	35	36	37
10	UK	32	33	35	36	36	36	36	–	–
	Anglo	32	33	34	35	36	38	40	40	40
	Asian	24	25	26	27	29	30	31	32	33
	Hisp	24	25	26	27	28	29	30	31	32
	Black	25	26	27	28	29	30	31	32	33
5	UK	27	28	29	30	33	33	33	–	–
	Anglo	27	28	29	30	32	34	36	37	38
	Asian	17	18	19	20	23	25	26	28	29
	Hisp	20	21	22	23	23	23	24	25	26
	Black	12	15	17	19	21	23	25	26	26
n (unweighted)	UK	174	185	180	196	189	191	171	–	–
	Anglo	46	59	44	52	53	36	56	40	49
	Asian	31	42	47	48	48	38	55	27	55
	Hisp	35	44	52	45	52	35	48	34	45
	Black	37	57	54	53	39	45	48	42	47

Note. The town name "Westown" was chosen, at the request of the school district, to preserve anonymity.

Studies with The Mill Hill Vocabulary Scale

So far, we have considered studies with the RPM. We turn now to the MHV.

The 1979 British Standardisation

The *Mill Hill Vocabulary Scale* (MHV) was standardised alongside the SPM in the 1979 study among young people, the sampling for which has already been described. As was the case with the SPM, there was no variance in MHV scores with region once the effect of Socio-Economic Status (SES) was partialled out (See Raven 1981; 2000b). SES explained 16.2% of the non-age-explained variance. MHV scores are, therefore, more related to background SES than SPM scores. Age accounted for 58% of the MHV variance. MHV scores did not plateau in adolescence in the same way as SPM scores; growth continued at approximately one and a half words per six-month interval through to age 15 ½ years.

As with the RPM, separate item analyses were carried out within eight SES groups. The reproducibility of the Scale properties across groups was again very high, averaging .97. The order in which children acquire knowledge of the meaning of words is therefore no more (and no less) affected by home background than is their ability to solve matrix problems. It would appear to be untrue that children from different backgrounds learn different subsets of dictionary words.

Table 8.4. *Standard Progressive Matrices*
US 1986 Data for Adolescents in Westown
Contributions of Ethnicity and Socio Economic Status to Total Variance

	Simple R	Mult. R	R sq.	R sq. change	Beta	Beta sq.
Age	29	29	8	8	27	7
Father's SES	-31	41	16	8	-20	4
Black	24	46	21	5	26	7
Hispanic	14	48	23	2	15	2
Asian	-04	48	23	0	0	0
Age	29	29	8	8	27	7
Black	24	38	14	6	26	7
Hispanic	14	44	19	5	15	2

Table 8.5. *Standard Progressive Matrices*
Correlations Between Item Difficulties Calculated Separately Within Specified Groups (Decimal point omitted and rounded to two decimal places)

		Westown Black	Westown White	Westown Hispanic	Westown Asian	Westown All	Des Moines	China
Westown	Black							
	White	98						
	Hispanic	100	98					
	Asian	98	99	98				
	All	99	99	100	99			
Des Moines		99	97	99	97	99		
China		95	94	94	96	95	96	
UK 1979		99	97	99	98	99	99	97

Standardisations in the US in the mid 80s

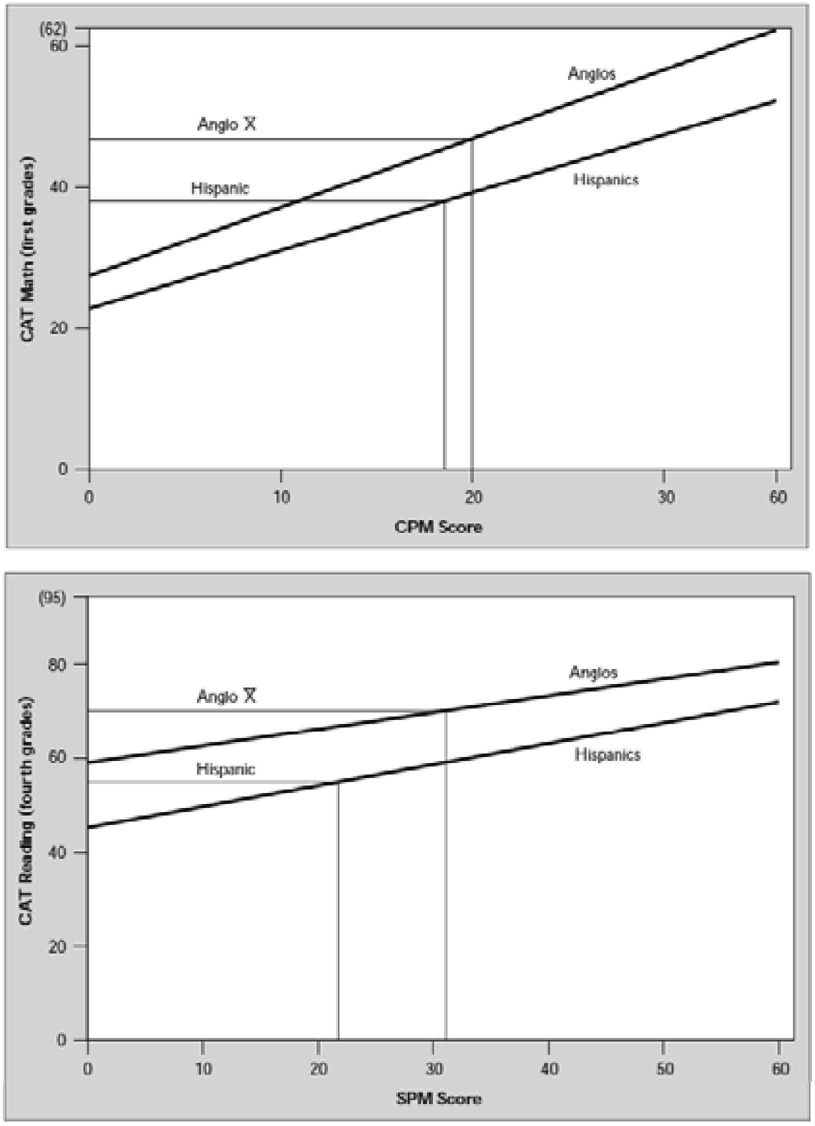
Many of the U.S. school districts that collected norms for the RPM did not administer the MHV. Nevertheless, as can be seen from Table 8.6, the overall U.S. norms for schoolchildren calculated from the data that were accumulated again lagged behind the international figures. However, the U.S. White norms once more corresponded fairly closely to those available for other cultures. As with the SPM, and as can be seen from Table 8.7, the test scaled in much the same way for (English speaking) students from different socioeconomic and ethnic backgrounds: Thus students from some backgrounds do *not* learn many of the kinds of word included in the Scale that are unknown to other cultural groups.

Studies of Adults

Standardisation in the U.K. in the Mid 1940s

The U.K. adult norms for the SPM that were published in the late 1940s and which formed the main reference data used worldwide for more than half a century were derived from a number of studies conducted between 1939 and 1947. Each of these samples and the way in which the resulting data were consolidated is described in some detail in Raven (2000b). Figure 8.3 shows the resulting norms in the form of a graph. Although the data were in fact collected over a number of years, they will, for convenience, hereinafter be referred to as the "1942 U.K. adult

Figure 8.2. *Coloured and Standard Progressive Matrices*
Sample regressions of the sub-tests of California Achievement Test on the RPM for Anglos and Hispanics in Douglas, Arizona.



Note: The upper figure shows the regressions of CAT Mathematics scores on the Coloured Progressive Matrices among first grade students. The lower figure shows the regressions of CAT Reading scores on the Standard Progressive Matrices among fourth grade students. (Redrawn from Hoffman, 1990.)

Table 8.6. *Mill Hill Vocabulary Scale*
1986 Adolescent Percentile Norms for the U.S.A. in the Context of 1979 British Data (Smoothed)

	Age in years (months)																			
	6½ 6(3) to 6(8)		7 6(9) to 7(2)		7½ 7(3) to 7(8)		8 7(9) to 8(2)		8½ 8(3) to 8(8)		9 8(9) to 9(2)		9½ 9(3) to 9(8)		10 9(9) to 10(2)		10½ 10(3) to 10(8)		11 10(9) to 11(2)	
Percentile	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**
95	23	**22**	24	**23**	26	**25**	28	**27**	30	**29**	32	**31**	34	**33**	36	**35**	38	**37**	40	**39**
90	21	**20**	22	**21**	24	**23**	26	**25**	28	**27**	30	**29**	32	**31**	34	**33**	35	**34**	37	**36**
75	18	**17**	19	**18**	20	**19**	22	**20**	24	**22**	26	**24**	28	**26**	30	**28**	32	**30**	34	**32**
50	**13**	**12**	**14**	**13**	**15**	**14**	**18**	**16**	**20**	**18**	**22**	**20**	**24**	**22**	**26**	**23**	**27**	**25**	**29**	**27**
25	8	**7**	9	**8**	10	**10**	13	**12**	15	**14**	17	**16**	19	**18**	20	**19**	22	**20**	24	**21**
10	6	**7**	6	**5**	7	**6**	9	**7**	10	**9**	11	**10**	13	**12**	15	**14**	18	**15**	20	**17**
5	5	**5**	5	**5**	5	**5**	6	**5**	7	**6**	7	**7**	9	**8**	11	**10**	13	**11**	15	**13**
n	103		135		149		175		157		183		205		179		209		200	

	11½ 11(3) to 11(8)		12 11(9) to 12(2)		12½ 12(3) to 12(8)		13 12(9) to 13(2)		13½ 13(3) to 13(8)		14 13(9) to 14(2)		14½ 14(3) to 14(8)		15 14(9) to 15(2)		15½ 15(3) to 15(8)		16 15(9) to 16(2)	16½ 16(3) to 16(8)
Percentile	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**	US	**US**
95	42	**41**	44	**43**	45	**45**	46	**47**	48	**49**	52	**51**	54	**53**	56	**55**	57	**56**	57	**57**
90	39	**38**	41	**40**	43	**42**	45	**44**	47	**46**	50	**48**	52	**50**	53	**52**	54	**53**	53	**54**
75	35	**33**	37	**34**	38	**36**	40	**38**	42	**40**	44	**42**	46	**44**	47	**45**	49	**46**	47	**48**
50	**31**	**28**	**32**	**30**	**33**	**32**	**35**	**33**	**36**	**34**	**38**	**36**	**40**	**38**	**41**	**39**	**43**	**40**	**42**	**43**
25	25	**23**	27	**25**	28	**26**	30	**28**	32	**30**	34	**32**	36	**33**	36	**34**	37	**35**	36	**36**
10	21	**18**	22	**19**	23	**21**	25	**23**	27	**24**	29	**26**	30	**27**	31	**28**	32	**29**	30	**30**
5	16	**14**	17	**16**	19	**17**	21	**19**	24	**21**	26	**23**	28	**25**	28	**26**	29	**27**	28	**28**
n	167		173		179		192		195		201		198		197		185			

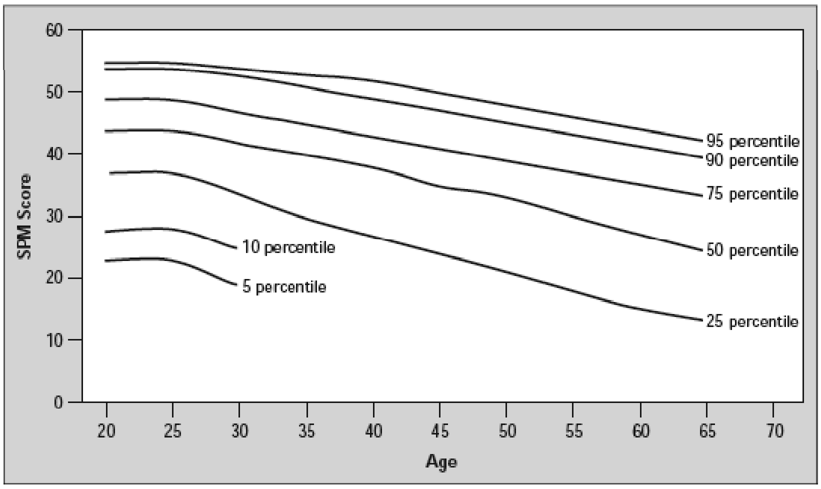
Note. US figures estimated on the basis of data available Summer 1986. The studies on which these norms are based are detailed in Raven et al. (1990). These show that the norms vary considerably between school districts, and, within districts, between ethnic groups.

norms". It will be seen that the scores obtained by successive age groups (after 20) become progressively lower. However, the scores of the less able appear to "decline" most. It must however, be stressed that the data for all age groups were, in essence, all collected at the same time. The graphs are *not* based on data collected from the same people as they got older. In other words they are based on cross-sectional as distinct from longitudinal data. As we shall see, interpreting the data *as if* they were longitudinal data has resulted in serious misunderstandings.

Table 8.7. *Mill Hill Vocabulary Scale*
Correlations Between Item Difficulties Calculated Separately Within Specified Groups (Decimal point omitted and rounded to two decimal places)

	Westown Black	Westown White	Westown Hispanic	Westown Asian
Westown Black				
Westown White	97			
Westown Hisp.	100	97		
Westown Asian	99	97	99	

Figure 8.3. *Standard Progressive Matrices*
The Apparent Decline in Scores as Age Increases as Documented in Typical Cross-Sectional Studies
UK Standardisation, circa 1942



Note: The graphs plot the cross-sectional norms for people of increasing age all tested, as described in Raven (2000a), in a series of studies conducted around 1942.

1992 Standardisation in Dumfries, Scotland

The first re-standardisation of the *Standard Progressive Matrices* in the U.K. (which was combined with the first ever standardisation of the *Advanced* Progressive Matrices on a general adult population) was carried out in Dumfries, Scotland, in 1992. Dumfries was chosen because: (a) the studies conducted among young people described earlier had made it clear that the most important variable to take into account in the choice of location was the balance of SES groups in the population, (b) these same studies had shown that the RPM norms obtained for the Borders region of Scotland – itself an area with a demographic structure which matches that of the U.K. as a whole – *did* correspond to those of the U.K. as a whole, and (c) the town of Dumfries recommended itself as a possible site for an adult standardisation because (i) much of the data collected with the RPM over the past 50 years (including two major standardisations of the *Coloured* Progressive Matrices) had been collected there and had stood up well in comparison with data collected elsewhere, (ii) it had itself a demographic structure which approximated that of the U.K., and (iii) it was geographically of such a size as to be easily traversed in search of named adults selected by systematic sampling procedures from a full list of adult residents. The procedures used for selecting, contacting, and testing the respondents, together with the response rates, are described in some detail in Raven (2000b).

1993 Standardisation in Des Moines, Iowa

Following the success of the Dumfries study, an exactly parallel study was carried out in Des Moines, Iowa. Des Moines is recognised as one of four U.S. cities having demographic compositions approximating the U.S. as a whole and is therefore widely used by researchers seeking a microcosm of America (*American Demographics,* May, 1985, pp. 38-42). While it is, of course, impossible for any one city – however closely its crude demographic statistics may correspond to the whole country – to match the whole at a detailed level, the studies conducted with the RPM among schoolchildren in the U.S. in the 1980s had in fact confirmed that the norms for Des Moines did approximate to those for the U.S. as a whole (J. Raven, 1989). Once again, the procedures used to draw the sample and contact potential respondents (together with response rates) are described in Raven (2000b).

Comparison of Dumfries and Des Moines Data

The norms for both the *Standard* and *Advanced* Progressive Matrices tests that were obtained in Dumfries and Des Moines were compared with each other. The APM norms for Des Moines are compared with those from Dumfries in Table 8.8. It can be seen that the Des Moines adult norms fall much where our previous research with the larger populations of schoolchildren would lead one to expect. The upper percentiles for Des Moines closely approximate those obtained in the U.K., while the 50th and lower percentiles – and especially the latter – lag behind, at least up to age 50. Nevertheless, our extensive research among young people (J. Raven, 2000a; J. Raven & Court, 1989) does suggest that the lower percentiles for the U.S. as a whole should lag further behind the U.K. norms than did those obtained in Des Moines. A number of possible explanations of this are discussed in J. Raven et al. (2000, updated 2004) but, whatever the explanation, the main point to be made here is that these Des Moines norms are probably above those which would have been obtained had a random sample of the entire U.S. population been tested.

Table 8.9 shows that the adult norms for the MHV for Dumfries and Des Moines were also similar.

The Effects of Date of Birth

It is time now to take another look at the variation in scores over time. To anticipate the outcome, we will find ourselves re-interpreting the data obtained from the previously mentioned cross-sectional studies which had been thought to show a significant decline in eductive ability, but less decline in reproductive ability, with increasing age. Our conclusion will be that a more accurate interpretation of these data is that most human abilities (including, for example, athletic ability), *but not reproductive ability* have improved rather dramatically over the past century and that this reveals a hitherto unsuspected effect of the environment on these scores. Yet the puzzle is not what it is usually taken to be. Since *most* abilities are improving, the question is why *reproductive* abilities show such little change despite investments in education and the mass media.

Eductive ability.

Figure 8.4 displays the 1979 normative data for the SPM which were derived from the previously described nationwide study of young people in

Table 8.8. *Advanced Progressive Matrices, Set II (Untimed)*
1993 Adult Percentile Norms for Des Moines, Iowa (U.S.) in the Context of 1992 Dumfries (U.K.) Data

	Age in years																					
	20		25		30		35		40		45		50		55		60		65		70	
	18		23		28		33		38		43		48		53		58		63		68+	
	to		to		to		to		to		to		to		to		to		to			
Perc-	22		27		32		37		42		47		52		57		62		67			
entile	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**
95	33	**32**	33	**32**	33	**32**	33	**32**	32	**32**	32	**32**	31	**31**	31	**30**	31	**29**	30	**27**	26	**25**
90	31	**30**	31	**30**	31	**30**	31	**30**	30	**30**	30	**30**	29	**29**	28	**28**	27	**27**	26	**25**	22	**23**
75	27	**27**	27	**27**	27	**27**	27	**26**	26	**26**	25	**26**	24	**26**	23	**25**	22	**24**	21	**22**	18	**19**
50	**22**	**20**	**22**	**20**	**22**	**20**	**21**	**19**	**20**	**19**	**19**	**19**	**18**	**19**	**17**	**18**	**16**	**16**	**15**	**14**	**13**	**12**
25	17	**15**	17	**15**	17	**15**	16	**15**	15	**15**	14	**14**	13	**14**	12	**13**	11	**12**	10	**10**	7	**8**
10	12	**10**	12	**10**	12	**10**	11	**10**	10	**10**	10	**10**	9	**9**	8	**8**	7	**7**	6	**6**	4	**4**
5	9	**7**	9	**7**	9	**7**	8	**7**	7	**7**	6	**7**	5	**6**	4	**5**	4	**4**	3	**3**	2	**1**
n	58	**28**	71	**53**	84	**72**	69	**77**	54	**121**	67	**69**	54	**33**	39	**36**	46	**27**	43	**33**	44	**54**

Note. Tests completed at leisure.

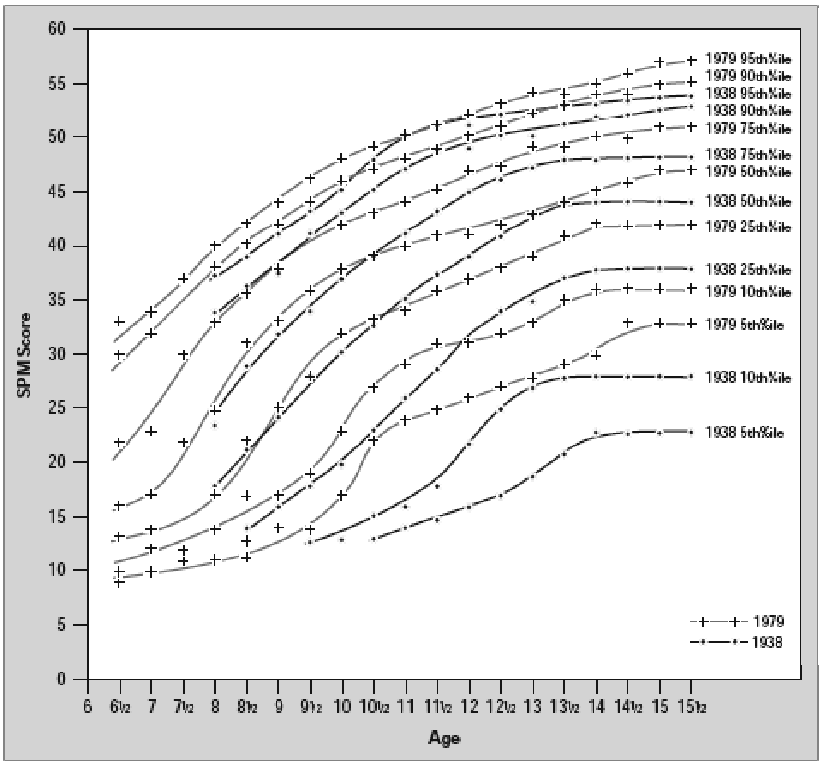
Table 8.9. *Mill Hill Vocabulary Scale, Forms 1 and 2 (Untimed)*
1993 Adult Percentile Norms for Des Moines, Iowa (U.S.) in the Context of 1992 Dumfries (U.K.) Data

	Age in Years																					
	20		25		30		35		40		45		50		55		60		65		70	
	18		23		28		33		38		43		48		53		58		63		68+	
	to		to		to		to		to		to		to		to		to		to			
Perc-	22		27		32		37		42		47		52		57		62		67			
	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**	UK	**US**
95	67	**69**	69	**70**	71	**71**	73	**72**	75	**73**	77	**74**	79	**75**	81	**76**	83	**77**	85	**78**	86	**77**
90	64	**63**	66	**65**	68	**66**	70	**67**	72	**68**	74	**69**	76	**70**	78	**72**	80	**73**	82	**73**	82	**72**
75	59	**56**	61	**57**	63	**58**	65	**59**	67	**61**	68	**63**	70	**65**	71	**66**	73	**68**	75	**68**	74	**67**
50	**53**	**51**	**55**	**52**	**57**	**53**	**58**	**54**	**60**	**55**	**61**	**57**	**62**	**58**	**63**	**60**	**64**	**62**	**65**	**62**	**63**	**61**
25	46	**44**	48	**46**	50	**47**	52	**48**	54	**50**	55	**51**	56	**52**	56	**53**	56	**53**	56	**53**	53	**52**
10	38	**36**	42	**38**	44	**40**	47	**42**	49	**44**	49	**46**	49	**46**	49	**46**	49	**46**	45	**46**	36	**45**
5	28	**23**	32	**25**	36	**27**	40	**31**	43	**43**	43	**35**	43	**37**	43	**38**	41	**38**	33	**38**	24	**36**
n	56	**26**	69	**53**	81	**70**	69	**75**	53	**118**	60	**68**	49	**31**	38	**35**	44	**29**	41	**32**	38	**56**

Note. Tests completed at leisure.

Great Britain (presented in Table 8.1) in the context of the data obtained in the 1938 Ipswich study. It is important to emphasise that the graphs within the chart do not show the scores obtained by the same young people as they grew older: They show the percentile scores obtained from a cross-section of young people of different ages who were tested in the same year. If one compares the graphs from the 1938 sample (i.e. the heavy lines) with those for the same percentiles for those tested in the 1979 sample (the light lines) it is clear that the level at which the scores plateau in adolescence has increased markedly and that young people

Figure 8.4. *Standard Progressive Matrices.*
Graphed percentile norms for young people in Great Britain in 1938 and 1979.



Note: The graphs show the score obtained by young people of different ages and levels of ability in these 2 years. If one compares the graph of the 1938 norms (i.e., the heavy lines) with those for the same percentile in 1979 (the light lines), it is clear that the level at which the scores plateau in adolescence has increased markedly and that young people get higher scores at earlier ages. (Thus, in the case of the 5th percentile, 10 1/2-year-olds in 1979 obtained similar scores to those obtained by 14-year-olds in 1938) get higher scores at earlier ages. (Thus, in the case of the 5th percentile, 10 ½ year olds in 1979 obtained similar scores to those obtained by 14 year olds in 1938.)

Martinolli (1990) with the CPM and Spicher (1993) with the SPM have demonstrated similar changes in the norms over time in Fribourg, Switzerland. Case and her colleagues (see Raven et. al. 2000, updated 2004) have likewise documented similar changes for Argentina from 1964 to 2000).

What these results do not show is whether the increase has been continuous and incremental or whether it occurred at a particular time, such as during the second world war.

Bouvier's (1969) data, derived from testing conscripts to the Belgian army each year from 1958 to 1967, and reproduced in Figure 8.5, suggest that the increase was steady rather than associated with any particular developments.

The SPM results from the 1992 adult standardisation in Dumfries are shown, plotted by date of birth, by the dotted lines in Figure 8.6. The dashed lines re-plot the scores obtained in the 1940s study previously shown in Figure 8.3.

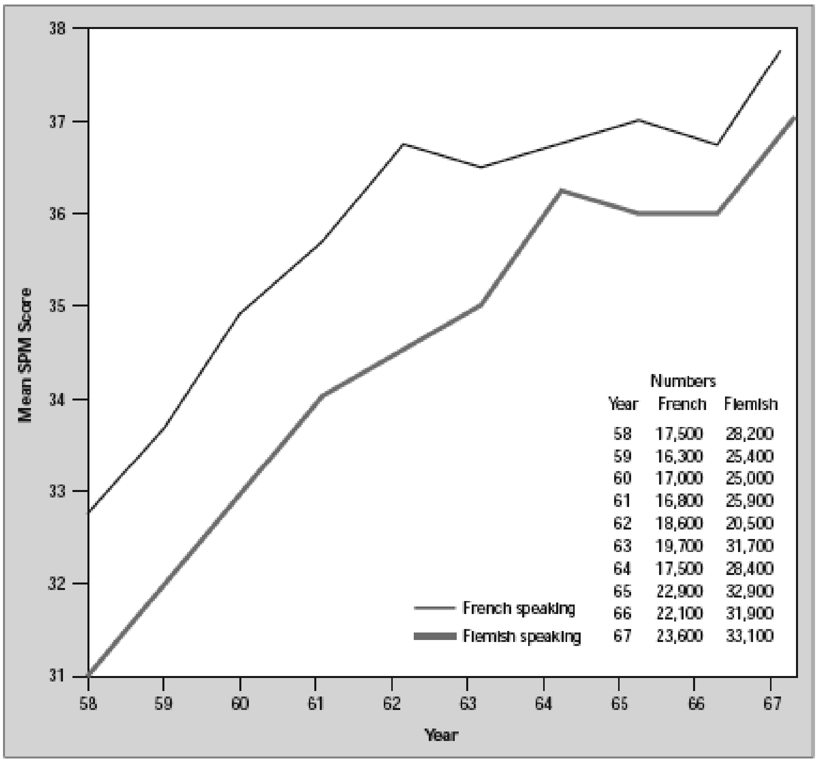
Examination of the points at which the two sets of graphs interface (i.e. among people born in 1922, where the earlier data are particularly strong) reveals that both the mean and spread of scores were very similar regardless of whether they were derived from the sample tested in the early 1940s (when they were roughly 20 years old) or from the sample tested in Dumfries in 1992 (when they were 70 years old). Instead of showing a decline in scores with advancing age (which is what the data behind each of these sets of graphs – and other similar data – had previously been thought to demonstrate), what the Figure clearly shows is a regular and continuous increase in the scores obtained by people born in different years, with the scores of the younger and more able respondents being the maximum obtainable on the test.

As previously noted, the continuity in the graphs derived from the two samples tested under different conditions in different places lends confidence to the adequacy of the data obtained in both studies.

This confidence is reinforced when data from a third – smaller – study conducted by Heron and Chown (1967) approximately halfway between the two studies already mentioned are introduced. The data have been superimposed on the graphs shown in Figure 8.6 in Figure 8.7. The graphs for the Heron and Chown data run straight through the point of interface between the graphs for the 1942 and 1992 data in the previous Figure. They thus confirm the adequacy of the data from both the previously mentioned studies.

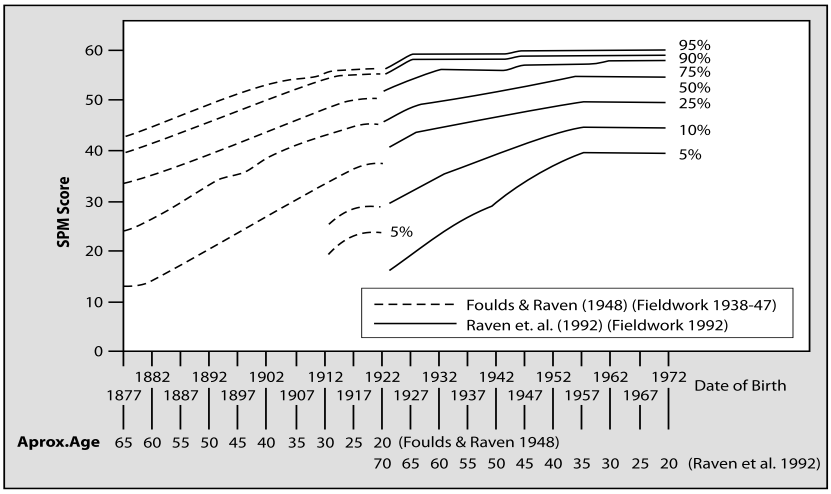
In an effort to guard against misleading conclusions being drawn from Figures 8.6 and 8.7, and because these Figures at first sight seem to confirm Teasdale and Owen's (1989) claim to "find no evidence of gains at the higher levels" it is important to note that the relatively small increase in scores among more able people born between 1922 and 1972 stems entirely from a ceiling effect on the SPM, which has only 60 items. As already mentioned, in both the Dumfries and Des Moines standardisations, the *Advanced* Progressive Matrices was standardised alongside the SPM. The APM norms for Dumfries for 1992 are compared with the 1962 norms for same test in Table 8.10. It is immediately

Figure 8.5. *Standard Progressive Matrices*
Mean scores of Belgian military conscripts from 1958 to 1967.



Note: The mean scores of French and Flemish-speaking recruits are graphed separately. (Redrawn from Bouvier, 1969, and reprinted with permission.)

Figure 8.6. *Standard Progressive Matrices*
100 years of eductive ability in Great Britain.
Graphed Percentile norms from the 1942 and 1992 standardisations plotted by date of birth.



Note: The Figure graphs the percentile norms obtained by adults of different ages (and thus dates of birth) on the Standard Progressive Matrices when a sample was tested circa 1942 (see comment on Figure 8.3) in one case and in 1992 in the other. The approximate age of people born in different years in the two samples is shown below.

Figure 8.7. *Standard Progressive Matrices*
UK standardisations, circa 1942, 1962 and 1992
100 years of eductive ability, including Heron & Chown's data from 1962.

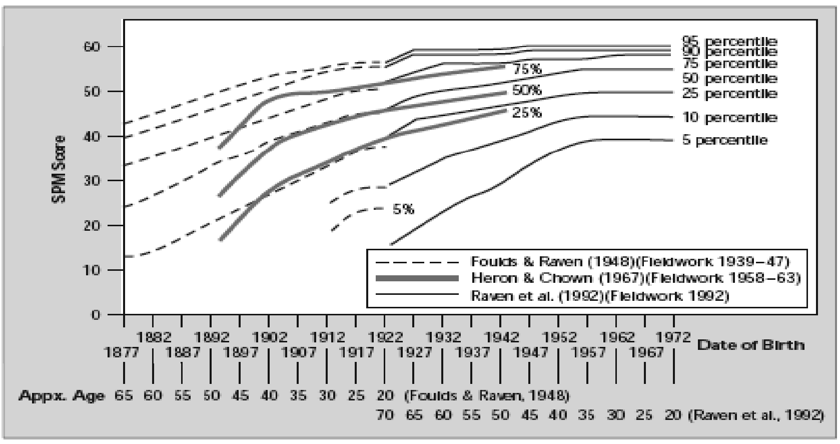


Table 8.10. *Advanced Progressive Matrices, Set II*
Comparison of 1992 and 1962 British Adult Percentile Norms

	Age in years					
	20		30		40	
Percentile	1962	**1992**	1962	**1992**	1962	**1992**
95	24	**33**	23	**33**	21	**32**
90	21	**31**	20	**31**	17	**30**
75	14	**27**	12	**27**	9	**26**
50	**9**	**22**	**7**	**22**	–	**20**

Note. The 1962 data (previously published in J. C. Raven, 1965) were estimated from the work of Foulds and Forbes, which was also published in J. C. Raven (1965). Since the test has 36 items and 8 options per item, scores of 6 or less verge on the chance level. There was therefore no point in publishing the lower percentiles in 1965.

obvious that the increase in scores evident in the lower percentiles in Figure 8.6 *has* been accompanied by major gains among the more able. (The enormous methodological difficulties which inhere in any attempt to isolate the relative size of gains at different points even on scales which conform to the Rasch model are discussed in the chapter by Prieler and Raven in this volume.) The effect was so great that the APM, which was originally developed to discriminate among the top 20% of the population, now offers an almost perfect Gaussian distribution across the entire adult population. Just as the entire distributions of height and athletic ability have moved up (with admittedly some change in shape), so has the entire distribution of eductive ability.

Reproductive ability

Turning now to reproductive ability, Table 8.11 compares the 1979 U.K. norms for adolescents on the *Mill Hill Vocabulary Scale* (MHV) with those obtained using the written test in Colchester in 1943. The 95th percentile has unmistakably dropped from 1943 to 1979. So has the 90th. The 75th has dropped, but the drop is less marked. The 50th is, to all intents and purposes, unchanged. The 25th had gone up. And the 10th and 5th show a marked increase.

Unfortunately, these apparently unambiguous results are not entirely confirmed when a comparison is made between the 1979 results and those obtained by oral administration of the MHV in 1943. Perhaps most importantly, whereas the comparison of the results obtained with the written test suggest a reduced variance in 1979, the comparison between the written test in 1979 and the oral test in 1943 indicate *increased*

Table 8.11. *Mill Hill Vocabulary Scale: Forms 1 and 2 (Self-Completed in Writing)*
UK data Norms for Adolescents from the 1979 Standardisation in the Context of 1943 Colchester Data

	Age													
	11½		12		12½		13		13½		14		15	
Percentile	43	**79**	43	**79**	43	**79**	43	**79**	43	**79**	43	**79**	43	**79**
95	41	**42**	47	**44**	50	**45**	52	**46**	54	**48**	57	**52**	60	**56**
90	40	**39**	43	**41**	47	**43**	49	**45**	51	**47**	53	**50**	55	**53**
75	34	**35**	36	**37**	40	**38**	43	**40**	44	**42**	45	**44**	48	**47**
50	**29**	**31**	**31**	**32**	**33**	**33**	**35**	**35**	**37**	**36**	**38**	**38**	**40**	**41**
25	24	**25**	26	**27**	27	**28**	29	**30**	30	**32**	31	**34**	33	**36**
10	17	**21**	19	**22**	21	**23**	22	**25**	24	**27**	25	**29**	26	**31**
5	12	**16**	14	**17**	16	**19**	17	**21**	18	**24**	19	**26**	20	**28**

Note. Based on samples of 1,419 (1943 data) and 1,304 (1979 data).

variance, with more able pupils appearing to know still more and less able pupils knowing still less!

Figure 8.8 presents the U.K. adult data. The graphs plot the percentile *Mill Hill Vocabulary Scale* scores achieved by a cross-section of adults tested during the 1940s alongside those obtained by a cross section of the population in 1992. It will be seen that, although there appear to have been some changes – with the scores obtained by less able middle-aged adults seeming to have gone up most – the changes are nothing like as great as those which have occurred in eductive ability.

Bouvier's (1969) data (Figure 8.9) likewise reveal little change in vocabulary test scores over the period of his study, especially among the French speaking group.

All these results suggest that reproductive ability – at least as assessed from people's knowledge of words – has changed much less than might have been expected, and certainly a great deal less than eductive ability, over the period for which data are available.

Schaie (1983, 1994) and Thorndike (1977) likewise concluded that it is the reasoning components of "intelligence" which have been increasing most rapidly and consistently. Their data are particularly interesting in that they show that this is true whether "reasoning ability" is measured by verbal or nonverbal tests and whether reproductive ability is measured by vocabulary or other routine skills like word fluency. On the other hand, their data do suggest that knowledge of vocabulary has increased rather more than the above data would lead one to expect and that scores on tests which require these two abilities to different

Figure 8.8. *Stability and Change in Reproductive Ability Over Time*
Mill Hill Vocabulary Scale: *Forms 1 and II*
Graphed Norms from Cross Sectional Norming Studies
Conducted among British Adults circa 1945 and 1992

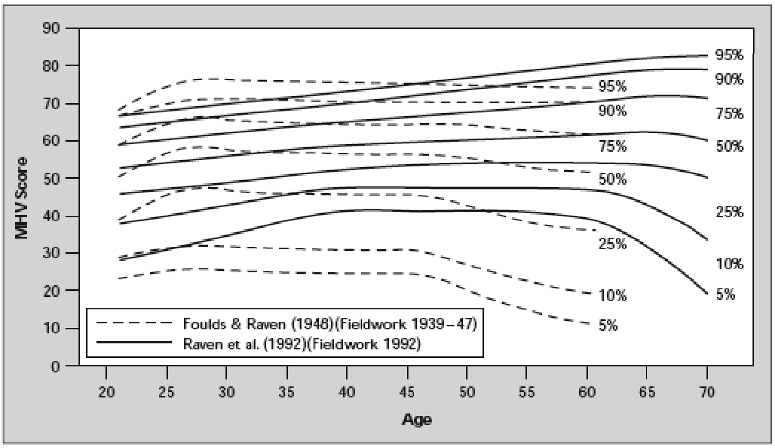
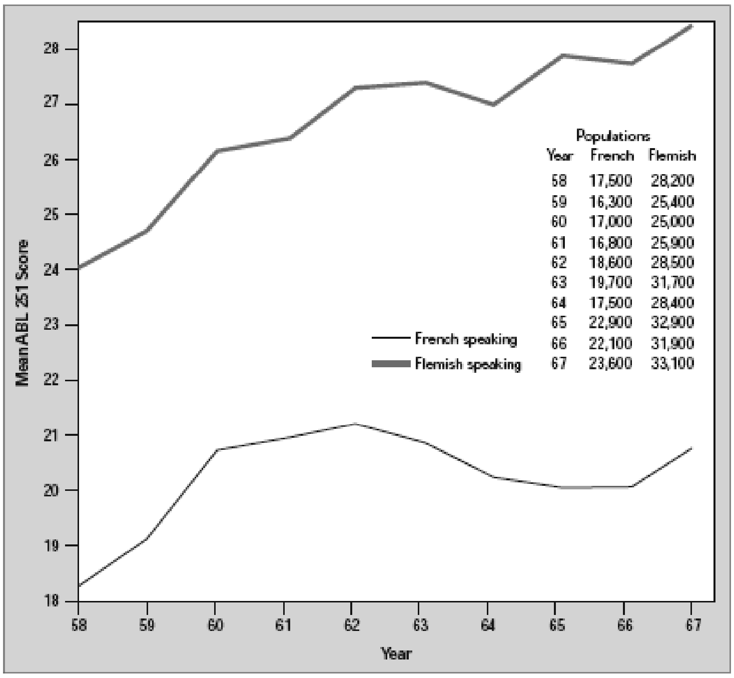


Figure 8.9. *Belgian Army Vocabulary Test (ABL 251)*
Mean scores of conscripts to the Belgian Army from 1958 to 1967.



Note: The mean scores of French and Flemish-speaking recruits are graphed separately. (Redrawn from Bouvier, 1969, and reprinted with permission.)

extents have increased in proportion to the extent to which they involve eductive ability. This inference has been strongly confirmed in Flynn's (2000) studies of the sub-scales of the WISC. The scores on these sub-tests have gone up in direct proportion to their loadings on an eductive ability factor. Schaie's position on these issues seems to have shifted over the years. In one publication (Schaie, 1983), he showed reasoning ability, whether measured by verbal or nonverbal tests increasing most, and numeric ability increasing and then declining, with other abilities falling in between. In a more recent article (Schaie, 1994), he presents graphs for what he calls "cohort gradients" for "latent abilities". According to these data, mean scores on "inductive reasoning" and "verbal memory" increased most steeply over the years. "Numeric ability" at first improved and then declined. His graph for what he calls "verbal ability" behaves somewhat similarly, but shows a later peak.

Discussion

It would appear from the results summarised above that there has been, and still is, considerable – if far from perfect – similarity in the SPM norms obtained in different societies with a tradition of literacy at any given point in time. However, in common with the scores on other tests, and especially those measuring eductive ability through verbal or nonverbal items – see for example Bouvier (1969), Thorndike (1975, 1977), Garfinkel and Thorndike (1976), and the large number of published and unpublished studies brought together by Flynn (1984, 1987) – there has been a continuous increase in the scores at all levels of ability over time.

The data on changes over time and the differences between ethnic groups naturally raise the question of what is responsible for these changes and differences. No one study – let alone any study of a correlational nature – can give a definite answer. But, since they do seem to make some hypotheses less likely and others more likely, it is worth summarising some of the data which bear on the question.

In what follows, the causes of the changes over time and the differences between ethnic groups will be considered simultaneously. On the one hand, the absence of cross-cultural differences in RPM scores between cultures which do differ on a variable which has been put forward as possible explanations of the changes over time make those explanations of the time differences less likely. On the other hand, variation in scores between cultural groups which do differ on a variable which has also changed over time and been suggested as a possible explanation of the time differences strengthens the possibility of that variable playing a significant role in the process.

Thorndike (1977) and Garfinkel and Thorndike (1976) listed a number of possible explanations of the time trends. However, the data available on the *Progressive Matrices* do not really support any them. Thorndike suggested for example, that the acceleration in development may be due to earlier maturity. However, if maturity is a factor, the curves plotting the age norms for boys and girls separately should differ more than the data published in J. Raven (1981) and J. Raven et al. (2000a) shows that they do. These data show that, with the exception of an unexplained divergence between the two curves at age 11 (when there is a school change) the curves are virtually identical. Furthermore that divergence itself has not been confirmed when we have plotted similar graphs for, e.g., a range of U.S. school districts. Likewise, he suggested that the increases may have been due to changes in the nature of early school education, but the fact that there was little difference between the RPM norms obtained in Scotland and England in the 1979 standardisation suggests that this is unlikely – because Scottish infant education remains very formal (HMI, 1980). The minor difference between the Chinese and British norms likewise tends to disconfirm this contention. Indeed, some of the school systems for which norms are available do not admit children until they are eight years old, and, as Thorndike himself noted, the largest increases seem to have occurred among children of preschool age. Thorndike suggests that television may have had an effect. However television was widely available in Ireland when what can now be seen to be low Irish norms were collected. Greenfield's (1998) argument that the change is due to familiarity with icons and computer games likewise does not hold up because, as Schaie (1983) has shown, there has been a *huge* increase in scores on *verbal* measures of "reasoning" (or eductive) ability.

Others have suggested that the increases in RPM scores over time may be attributed to schools using matrix-type problems to teach "problem solving". However, Thorndike showed that performance on *all* the subscales of the Stanford-Binet had improved and that the greatest increases were among very young children who had not yet started school. In our own data there is little difference between the norms from cultures which differ markedly in the age at which children start school.

Flynn (having, in 1984, queried Thorndike's hypotheses concerning the Binet results) likewise concluded in his 1987 article that most of the common and obvious explanations of the RPM increase do not hold up. Among other things he showed, through a detailed analysis of de Leeuw and Meester's (1984) data, that changes in the amount of education people have could account for only one of the 20 IQ-point gain in RPM scores documented among servicemen. Changes in the intellectual quality of the home environment – at least insofar as it is indexed by SES – could account for little more.

In summary, then, most of the common explanations of the changes over time do not hold up: Where there is variation between cultures in a variable which potentially help to explain the change over time it is not accompanied by differences in RPM scores. Having, in this way, made such explanations *less* likely (although not, of course, ruling them out) it behoves us to look elsewhere.

A potentially more fruitful line of enquiry is suggested by the fact that the variation in mean scores between ethnic groups within the U.S. does seem to correspond to variation between the same groups in height, birth weight, and infant mortality. Height and birth weight have, like intelligence test scores, increased over the past 80 years (Floud, Wachter, & Gregory, 1990). These observations led us to suspect that the increase in RPM scores over time might be attributable to the same factors as have been responsible for increases in height and birth weight and for decline in infant mortality – that is, to improved nutrition, welfare, and hygiene. This statement does not, unfortunately, get us very far since what it is about the causes of these changes remains obscure. But then, so do the ways in which most drugs produce their effects: The fact that one cannot offer a complete explanation does not undermine the value of what one *can* do to clarify the position.

On looking for literature which bore on the nutritional hypothesis we unearthed a remarkable, although largely unpublished, study carried out in Aberdeen, Scotland, which showed that dietary and hygiene variables *do* influence RPM scores as well as birth weight and height (Baird & Scott, 1953; Scott, Illsley, & Thomson, 1956). In this study, calcium intake was used as an index of quality of diet. This had a marked impact on all three of the outcomes mentioned – and the relationship held up both within and between socioeconomic groups. Vernon (1969) had come to a similar conclusion.

More recently, Benton and Roberts (1988), Benton and Butts (1990), Benton and Cook (1991), Nidich, Morehead, Nidich, Sands, and Sharma (1993), Schoenthaler, Amos, Doraz, Kelly, and Wakefield (1991), Schoenthaler, Amos, Eysenck, Peritz, and Yudkin (1991), and others have shown that vitamin and mineral supplements have a rapid and marked effect on some people's eductive, but not reproductive, ability scores.

Confirmation of the nutrition/hygiene hypothesis comes from the reviews by Pollitt and Saco-Pollitt (1996) and Sigman and Whaley (1998) of physical factors affecting intelligence. Low levels of iodine in diet had a major effect in the U.S. well into this century, and still do in other countries. Intestinal parasites have similar effects. Low oxygen pressures arising from high altitude also have a marked effect (cf. the previously mentioned Peruvian mountain norms).

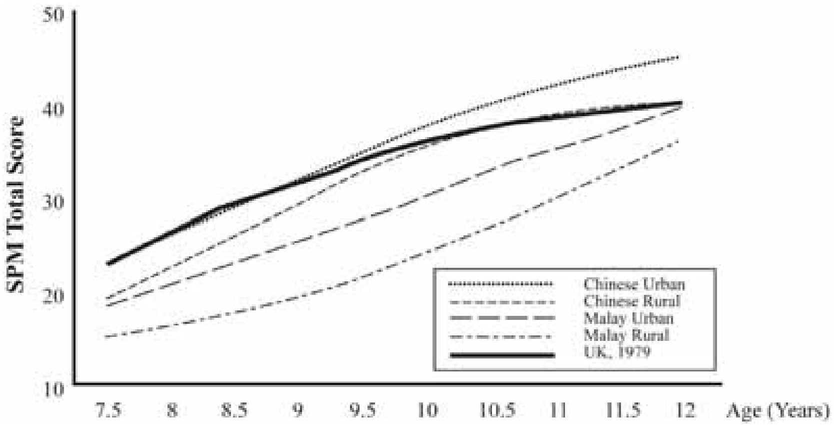
Although all these studies seem to support the contention that the differences over time and between cultural groups reflect the balance of vitamins available in the diet (*not*, pace Martorell, 1998, the absolute quantity of food) with isolated mountainous areas – which are unlikely to have extensive trade in agricultural products – having the lowest scores, this conclusion does not receive unequivocal support in the literature.

Chiam's data is particularly disconcerting. In two senses. First, in order to determine what they had been eating, Chiam (1995) collected samples of faeces from a cross-section of children for whom a variety of test scores (but not RPM scores) were available. She found that diet was unrelated to any of them. In another study (Chiam, 1994) of 5,412 children aged 7 ½ to 12 years in contrasting areas of Malaysia, she found that, as Figure 8.10 shows, the RPM norms for the Chinese in Malaysia corresponded to the international norms while those for the Malay population did not. If diet were responsible for the differences, one would expect them to be reduced when such things as the effects of urban and rural residence and SES were partialled out. Yet, although, as can be seen, there were differences between the urban and rural norms for both groups, this variable did not account for the differences between the groups. Neither difference was accounted for by education or socioeconomic status. Once again, then, the analyses render less likely the most commonly offered explanations of group differences and changes over time but in this case also throw doubt on the dietary hypothesis.

Some implications

It seems from this review of available literature that most of the most commonly offered explanations of the changes over time (including those promoted by such authors as Greenfield & Williams in Neisser's 1998,

Figure 8.10. *Standard Progressive Matrices*
Mean Scores For Young People in Malaysia by Ethnicity and Area of Residence
1992 Data



From Chiam (1994)

book) are unlikely to be true. Even the most likely hypothesis – the diet/welfare/hygiene hypothesis – does not receive unequivocal support. But whatever their *causes*, the changes which have been documented point to a clear, major, and previously unsuspected effect *of the environment* on eductive ability.

It follows from this that, despite their persistence over time as the scores of *all* groups have been going up (see also the persistence of the French-Flemish speaking difference in Belgium shown in Figures 8.5 and 8.9), the differences between ethnic groups within the U.S. cannot be regarded as immutable. It was, of course, precisely *this* point – and not some more general argument about the nature of "intelligence" – that Flynn sought to establish when he set about documenting and publicising the mutability of "intelligence" test scores. The point may be underlined by noting that most of the RPM norms for ethnic norms within the U.S. currently lie between the 1938 and 1979 British norms, that is, within the range over which environmentally induced change might not only have been theoretically possible but has actually been demonstrated.

Other features of the environment which make a difference

Although the effects are insufficient to explain the gross time and cultural differences discussed in this article, and although it would not be appropriate to present a thorough review of the relevant literature here, there have been a number of empirical studies of factors which increase or decrease RPM scores and it is worth mentioning some of them as a counterpoint to simplistic hereditatrian and dysgenic arguments about "intelligence". The results surprise many psychologists. Eductive ability has turned out to be more easily influenced by appropriate educational and developmental experience than reproductive ability. However, the variables which influence the development of eductive ability are *not* the obvious cultural and socioeconomic variables which divide society and on which sociologists have focused so much attention. Acquired information *is* more influenced by these variables than is the ability to perceive and think clearly – but these background variables still account for only a small proportion of the total variance.

Eductive Ability, Upbringing, and Education

Many studies (e.g. Chan, 1981; McGillicuddy-DeLisi, 1985; McGillicuddy-DeLisi, DeLisi, Flaugher, & Sigel, 1987; J. Raven, 1980; Sigel & Kelley, 1988) have shown that the development of children's eductive ability is promoted if their parents involve them in their own thought processes. Such parents involve their children in their own attempts to make sense of difficult situations, as they use their feelings as a basis for "experimental" action, as they resolve value conflicts, and as they consider the long-term social consequences of their actions. All this necessitates that parents share with their children their own understanding of the workings of society and their role in it. The children are thereby presented with a thought process which is fundamentally conceptual, yet which also relates thought to action. Such parents are also more likely to treat their children with respect, and realise the need to earn (rather than demand) their children's respect. This leads them to initiate a cyclical process in which they discover just how competent their children really are and, as a result, become more willing to place them in situations which call for high-level competencies. The result is that their children have many opportunities to practice and develop these competencies. Such parents are more inclined to read to their children stories which bear on moral problems. The outcome is that the children empathise with the various characters in the books and are able to reach their own moral position. The importance of reading *to* children in the development of their moral character and analogical reasoning has been underlined in the work of Jackson (1986) and Vitz (1990).

J. Raven (1980, 1987, 1989) and Vygotsky (1978, 1981) have shown that the above is only part of a wider process whereby parents who effectively nurture high-level competencies in their children tailor environments to the motives, incipient talents, and problems of their children. This is one way in which, as Plomin (1989) and Plomin and Daniels (1987) have shown, the within-family variance in children's environments becomes considerable and linked to variance in inherited characteristics in a way which markedly affects their development. As Scarr, Webber, Weinberg, and Wittig (1981) have noted, a similar effect is produced as children select themselves into different environments.

It follows from these observations that, if we wish to identify the genetic and environmental variables which influence psychological development, we will need to develop a more sophisticated model of the process.

The development of eductive ability in schools (but only in some cases measured by the RPM) has been studied by a number of researchers. Nickerson, Perkins, and Smith (1985) and Stallings and Kaskowitz (1974) found that the development of eductive ability is promoted by at least some forms of "open" or "progressive" education. Miller, Kohn, and Schooler (1985, 1986) and J. Raven, Johnstone, and Varley (1985) found that educational self-direction (i.e. pupils taking responsibility for their own education and moral decisions) and the undertaking of more complex educational activity (e.g. project-based, enquiry-oriented work) gave rise to a cyclical development in cognitive ability. Greater emphasis on self-direction and the development of new understanding fosters student competence, which in turn increases students' desire to gain more control over their destinies and encourages teachers' willingness to rely on their pupils' abilities.

Schooler, Mulatu, and Mesfin (2001), in the course of a 30-year follow up of a sample originally interviewed and tested in 1964, have confirmed their earlier longitudinal work (conducted mainly with Kohn) showing that substantively complex work improves intellectual functioning and, in a remarkable experimental study, Lovaglia, Lucas, Houser, Thye, and Markovsky (1998) have shown that even relatively minor, experimentally-induced, changes in perceived status produce significant (half-standard deviation) changes in RPM scores. (It may be worth noting that a change of this magnitude is greater than is typically achieved by training in the methods required to solve the problems.)

Having reviewed material demonstrating the importance of certain child development and educational practices in promoting the development of eductive ability, it is important to repeat that none of the psychological and educational processes mentioned above produce effects sufficient to account for the inter-generational increase in RPM scores. Furthermore, none of the activities described in the studies published to date significantly reduce the variance within socioeconomic groups and within families. Yet the within-family variance amounts to two thirds of the variance in test scores. It therefore seems that the environmental factors which have most influence on eductive ability are not the psychological and educational variables with which psychologists have been preoccupied in the past, and they appear to have little effect on its heritability.

The Quest for Single-Factor Explanations

In the *General Introduction* to this book, we argued that the evidence that both the RPM and MHV conformed to the requirements of Item Response Theory showed that *eductive* and *reproductive* abilities are every bit as "real" and measurable as the "hardness" of substances in geology, "height" in physiology, "high jumping ability" in athletics, and "life expectancy" in actuarial accounting.

In the same chapter we included a graph showing that life expectancy for men in the U.K. has almost doubled over the century for which RPM data have been available. We suggested that this graph undermined most of the arguments Flynn had put forward in his attempt to undermine the meaningfulness of psychological tests: He had argued, for example, that back projection of the increases in RPM and other "intelligence" test scores to the time of the Greeks would mean that they must have been incredibly stupid and, since this could not have been the case, that the tests must be meaningless. Unfortunately for Flynn's case, the same arguments would apply to height, athletic ability, and life expectancy. (As will be seen in later chapters, we do not dispute his argument that the importance of these psychological abilities as determinants of life performance has been exaggerated.)

We used the same analogies to draw attention to the illogicality in other arguments which crop in discussions of the nature of eductive and reproductive abilities and the environmental variables which affect them. No one would use the scaleability of any of the previously mentioned variables as a basis from which to argue that the observable and measured variance must be due to a single underlying cause in the way in which many psychologists have argued that, because eductive ability is scaleable, the variance must be due to a single underlying factor like "speed of neural transmission". Nor would they be tempted to argue that, because high jumping ability is clearly trainable, the measure is meaningless. Still less would they be inclined to set out to search for a single cause for the increases in height, high jumping ability, and life expectancy that have occurred over the years.

Yet there is one final, very intriguing, and almost completely overlooked question which emerges from this research. This is: Why is it that, when everything else has been going up … height, athletic ability, life expectancy … *reproductive* ability, whether measured by the MHV or any of a host of other measures that exist in this area, has been increasing hardly at all? This despite the massive investments that have been made in education, mass media, and information technology. In a sense, Flynn has focused the attention of psychologists on exactly the wrong question.

Concluding Asides

In concluding, it seems appropriate to draw attention to the seriousness of the errors which stem from the use of outdated norms. In the first place, it is obvious from Figure 8.6 that a score that would place a 50 year old tested in 1942 at the 95th percentile if judged against the 1942 norms would result in classification as at the 25th percentile if judged against today's norms. Such huge discrepancies in the interpretation of scores mean that the use of out-of-date norms cannot be justified: They are bad for the individuals concerned, bad for the organisations for which they work, and bad for society.

Still more serious, however, are the errors which arise from the adoption of out-of-date norms in research. The effectiveness of such things as educational enrichment programs is typically evaluated by comparing the scores obtained by experimental groups with published norms. When these norms are out-of-date, such experimental programs can only appear to be much more effective than they are.

References

Adams, E. A. (1952). *Analysis of Raven's Matrices Scores: Preliminary Report.* Surrey, England: Surrey Educational Research Association.

Angelini, A. L., Alves, I. C. B., Custodio, E. M., & Duarte, W. F. (1988). *Manual Matrizes Progressivas Coloridas.* Sao Paulo, Brazil: Casa do Psicologo.

Baird, D., & Scott, E. M. (1953). Intelligence and childbearing. *Eugenics Review, 45(3),* 139-154.

Benton, D., & Butts, J. (1990). Vitamin/mineral supplementation and intelligence. *The Lancet, 335*, 1158-1160.

Benton, D., & Cook, R. (1991). Vitamin and mineral supplements improve the intelligence scores and concentration of six-year-old children. *Personality and Individual Differences, 12*, 1151-1158.

Benton, D., & Roberts, G. (1988). Effect of vitamin and mineral supplementation on intelligence of a sample of schoolchildren. *The Lancet, 23*, January, 140-143.

Bouvier, U. (1969). *Evolution des Cotes a Quelques Test.* Belgium: Centre de Recherches, Forces Armees Belges.

Byrt, E., & Gill, P. E. (1973). *Standardisation of Raven's Standard Progressive Matrices and Mill Hill Vocabulary for the Irish Population: Ages 6-12.* Unpublished Master's Thesis, National University of Ireland, University College Cork.

Chan, J. (1981). Correlates of parent-child interaction and certain psychological variables among adolescents in Hong Kong. In J. L. M. Binnie-Dawson (Ed.), *Perspectives in Asian Cross-Cultural Psychology* (pp. 117-131). Lisle, Netherlands: Swets and Zeitlinger.

Chan, J. (1989). The use of Raven's Progressive Matrices in Hong Kong: A critical review. *Psychological Test Bulletin,* November, *2*(2), 40-45. Hawthorn, Victoria: ACER.

Chiam, H. K. (1994). *Is the Raven Progressive Matrices valid for Malaysians?* Paper presented to the 23rd International Congress of Applied Psychology, Madrid.

Chiam, H. K. (1995) *The standardisation of several tests in Malaysia.* Unpublished manuscript, School of Education, University of Malaya, Kuala Lumpur.

Court, J., & Raven, C. J. (2001). *A Researcher's Bibliography for Raven's Progressive Matrices and Mill Hill Vocabulary Scales.* Obtainable in hard copy and disk format from Susan Middleton, Harcourt Assessment, 19500 Bulverde Rd., San Antonio, Texas 78259, USA. <susan_middleton@harcourt.com>

Court, J. H., & Raven, J. (1995). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 7: Research and References: Summaries of Normative, Reliability, and Validity Studies and References to all Sections.* San Antonio, TX: Harcourt Assessment.

Dahlstrom, W. G. (1993). Tests: Small samples, large consequences. *American Psychologist, 48(4),* 393-399.

de Leeuw, J., & Meester, A. C. (1984). Over het intelligente-nderzoek bij de militaire keuringen vanaf 1925 tot henden. [Intelligence-as tested at selections for the military service from 1925 to the present]. *Mens en Maatschappij, 59,* 5-26.

de Lemos, M. M. (1984). A note on the Australian norms for the Standard Progressive Matrices. *Bulletin for Psychologists, (ACER Melbourne), 36,* 9-12.

de Lemos, M. M. (1989). The Australian re-standardisation of the Standard Progressive Matrices. *Psychological Test Bulletin,* November, *2*(2), 17-24. Hawthorn, Victoria: ACER.

Ferjencik, J. (1985). *Manual: Coloured Progressive Matrices.* Bratislava, Slovakia: Psychodiagnosticke a Didakticke Testy.

Floud, R., Wachter, K., & Gregory, A. (1990). *Height, Health, and History.* Cambridge, England: Cambridge University Press.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95,* 29-51.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101,* 171-191.

Flynn, J. R. (2000). IQ gains, WISC subtests and fluid ***g***: ***g*** theory and the relevance of Spearman's hypothesis to race. In G. R. Bock, J. A. Goode, & K. Webb (Eds.), *The Nature of Intelligence (Novartis Foundation Symposium 233)* pp. 202-227. Chichester, England: Wiley.

Garfinkel, R., & Thorndike, R .L. (1976). Binet item difficulty: Then and now. *Child Development, 47,* 959-965.

Greenfield, P .M. (1998). The cultural Evolution of IQ. In U. Neisser (Ed.), *The Rising Curve* (pp. 81-124). Washington, DC: American Psychological Association.

Heron, A., & Chown, S. (1967). *Age and Function.* London: Churchill.

HMI (Scotland) (1980). *Learning and Teaching in Primary 4 and Primary 7.* Edinburgh: HMSO.

Hoffman, H. V. (1983). *Regression Analysis of Test Bias in the Raven's Progressive Matrices for Anglos and Mexican-Americans.* Unpublished doctoral dissertation, Department of Educational Psychology, Graduate College, University of Arizona.

Hoffman, H. V. (1990). In J. Raven et al., A *Compendium of North American Normative and Validity Studies, Research Supplement No.3 to the Manual for Raven's Progressive Matrices and Vocabulary Tests* (pp. 21-31). San Antonio, TX: Harcourt Assessment.

Hollingshead, A. B. (1967). In C. M. Bonjean, R. J. Hill, & S. D. McLemore (Eds.), *Sociological Measurement: An Inventory of Scales and Indices.* San Francisco: Chandler.

Holmes, B. J. (1980). *British Columbia Norms for Wechsler Intelligence Scale for Children-Revised; Peabody Picture Vocabulary Test; Slosson Intelligence Test; Standard Progressive Matrices, and Mill Hill Vocabulary Scale.* Unpublished manuscript, University of British Columbia: Faculty of Education.

Hyman, H. H. (1955). *Survey Design and Analysis: Principles, Cases, and Procedures.* Glencoe, IL: Free Press.

Jackson, P. W. (1986). *The Practice of Teaching.* New York: Teachers College Press.

Jaworowska, A., & Szustrowa, T. (1991). *Podrecznik Do Testu Matryc Ravena.* Warsaw: Pracownia Testow Psychologicznych Polskiego Towarzystwa Psychologicznego.

Jensen, A. R. (1974). How biased are culture-loaded tests? *Genetic Psychology Monographs, 90,* 185-244.

Kahn, H., Spears, J., & Rivera, L. (1977). *Applicability of Raven Progressive Matrices Tests with School Children in Puerto Rico.* Hato Rey, Puerto Rico: Department of Education.

Kratzmeier, H., & Horn, R. (1979). *Manual: Raven-Matrizen-Test, Standard Progressive Matrices.* Weinheim, Germany: Beltz Test.

Lovaglia, M. J., Lucas, J. W., Houser, J. A., Thye, S. R., & Markovsky, B. (1998). Status processes and mental ability test scores. *American Journal of Sociology, 1,* 195-228.

Martinolli, L. (1990). *Etude et Reetalonnage des Matrices Progressives Couleur.* Unpublished manuscript, L'Institute de Psychologie, Universite de Fribourg.

Martorell, R. (1998). Nutrition and the worldwide rise in IQ scores. In U. Neisser (Ed.), *The Rising Curve*. Washington, DC: American Psychological Association.

McGillicuddy-DeLisi, A. V. (1985). The relationship between parental beliefs and children's cognitive level. In I. E. Sigel (Ed.), *Parental Belief Systems: The Psychological Consequences for Children* (pp. 261-299). Hillsdale, NJ: Erlbaum.

McGillicuddy-DeLisi, A. V., DeLisi, R., Flaugher, J., & Sigel, I. E. (1987). Family influences on planning. In S. L. Friedman, E. K. Scholnick, & R. R. Cocking (Eds.), *Blueprints for Thinking: The Role of Planning in Cognitive Development* (pp. 395-427). New York: Cambridge University Press.

Mehlhorn, H. G. (1980). Aspekte der Geistigen Entwicklung Jugendlicher. In W. Freidrich und H. Muller (Eds.), *Zur Psychologie der 12 bis 22 Jahrigen.* Berlin, Germany: VEB Deutscher Verlag der Wissenschaften.

Miao, E. S. Y. (1993). Translation of J. Raven, J. C. Raven, & J. H. Court, *Manual for Raven's Progressive Matrices Tests.* Taiwanese Edition. Taiwan: Chinese Behavioural Science Corporation.

Miao, E .S. Y., & Huang, W. (1990, July). *Standardisation and validation of CPM, SPM, and APM in Taiwan, ROC.* Paper presented to 22nd International Congress of Applied Psychology, Kyoto, Japan.

Miller, K. A., Kohn, M. L., & Schooler, C. (1985). Educational self-direction and the cognitive functioning of students. *Social Forces, 63,* 923-944.

Miller, K. A., Kohn, M. L., & Schooler, C. (1986). Educational self-direction and personality. *American Sociological Review, 51,* 372-390.

Neisser, U. (Ed.) (1998). *The Rising Curve.* Washington, DC: American Psychological Association.

New Zealand Council for Educational Research. (1984). *Standard Progressive Matrices: New Zealand Norms Supplement.* Wellington, New Zealand: NZCER.

Nidich, S. I., Morehead, P., Nidich, R. J., Sands, D., & Sharma, H. (1993). The effect of the Maharishi Student Rasayana Food Supplement on non-verbal intelligence. *Personality and Individual Differences, 15*, 599-602.

Nickerson, R., Perkins, D. N., & Smith, E. (1985). *The Teaching of Thinking.* Hillside, NJ: Erlbaum.

Owen, K. (1992). The suitability of Raven's Standard Progressive Matrices for various groups in South Africa. *Personality and Individual Differences, 13*, 149-159.

Owens, W. A. (1966). Age and mental abilities: A second adult follow-up. *Journal of Educational Psychology, 57,* 311-325.

Pollitt, E., & Saco-Pollitt, C. (1996). On the role of the physical environment in the development of intelligence. In D. K. Detterman (Ed.), *Current Topics in Human Intelligence*: *Volume 5*: *The Environment.* Norwood, NJ: Ablex.

Plomin, R. (1989). Environment and genes. *American Psychologist, 44*(2), 105-111.

Plomin, R., & Daniels, D. (1987). Why are children in the same family so different from one another? *Behavioral and Brain Sciences, 10,* 1-15.

Raven, J. (1980). *Parents, Teachers and Children*: *An Evaluation of an Educational Home Visiting Programme.* Edinburgh: Scottish Council for Research in Education.

Raven, J. (1981). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Research Supplement No.1, The 1979 British Standardisation of the Standard Progressive Matrices and Mill Hill Vocabulary Scales, Together With Comparative Data From Earlier Studies in the U.K., US, Canada, Germany, and Ireland.* San Antonio, TX: Harcourt Assessment.

Raven, J. (1987). Values, diversity, and cognitive development. *Teachers College Record, 89,* 21-38.

Raven, J. (1989). The Raven Progressive Matrices: A review of national norming studies and ethnic and socioeconomic variation within the United States. *Journal of Educational Measurement, 26,* 1-16.

Raven, J. (1995). Methodological problems with the 1992 standardisation of the SPM: A response. *Personality and Individual Differences, 18*(3), 443-445.

Raven, J. (2000a). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Research Supplement No.3 (Second Edition)*: *A Compendium of International and North American Normative and Validity Studies Together with a Review of the Use of the RPM in Neuropsychological Assessment by Court, Drebing, & Hughes*. San Antonio, TX: Harcourt Assessment.

Raven, J. (2000b). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology, 41*, 1-48.

Raven, J., Johnstone, J., & Varley, T. (1985). *Opening the Primary Classroom.* Edinburgh: Scottish Council for Research in Education.

Raven, J., Raven, J. C., & Court, J. H. (1998a, updated 2003). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 1*: *General Overview.* San Antonio, TX: Harcourt Assessment.

Raven, J., Raven, J. C., & Court, J. H. (1998b). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 2, The Coloured Progressive Matrices.* San Antonio, TX: Harcourt Assessment.

Raven, J., Raven, J. C., & Court, J. H. (1998c). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 4, The Advanced Progressive Matrices.* San Antonio, TX: Harcourt Assessment.

Raven, J., Raven, J. C., & Court, J. H. (1998d). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 5*: *The Mill Hill Vocabulary Scale.* San Antonio, TX: Harcourt Assessment.

Raven, J., Raven, J. C., & Court, J. H. (2000, updated 2004). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 3*: *The Standard Progressive Matrices*. San Antonio, TX: Harcourt Assessment.

Raven, J. C. (1941). Standardisation of Progressive Matrices, 1938. *British Journal of Medical Psychology, XIX,* Part 1, 137-150.

Raven, J. C., Court, J. H., & Raven, J. (1995). *Raven, Matrices Pogresivas (Escalas: CPM, SPM, APM).* Madrid: TEA Ediciones, S.A.

Raven, J. C., & Walshaw, J. B. (1944). Vocabulary tests. *British Journal of Medical Psychology, 20,* 185-194.

Sahin, N., & Duzen, E. (1994). *Turkish Standardization of the Raven's SPM (Ages 6-15).* Paper presented at the 23rd International Congress of Applied Psychology, 17-22 July, Madrid, Spain.

Scarr, S., Webber, P. L., Weinberg, R. A., & Wittig, M. A. (1981). Personality resemblance among adolescents and their parents in biologically related and adoptive families. *Journal of Personal Social Psychology, 40,* 885-898.

Schaie, K. W. (Ed.). (1983). *Longitudinal Studies of Adult Psychological Development.* New York: Guilford Press.

Schaie, K. W. (1994). The course of adult intellectual development. *American Psychologist, 49*(4), 304-313.

Schoenthaler, S. J., Amos, S. P., Doraz, W. E., Kelly, M. A., & Wakefield, J. (1991). Controlled trial of vitamin-mineral supplementation on intelligence and brain function. *Personality and Individual Differences, 12(4)*, 343-350.

Schoenthaler, S. J., Amos, S. P., Eysenck, H. J., Peritz, E., & Yudkin, J. (1991). Controlled trial of vitamin-mineral supplementation: Effects on intelligence and performance. *Personality and Individual Differences, 12(4)*, 351-362.

Schooler, C., Mulatu, M. S., & Mesfin, S. (2001). The continuing effects of substantively complex work on the intellectual functioning of older workers. *Psychology and Aging 16(3) 466-482*.

Scott, E. M., Illsley, R., & Thomson, A. M. (1956). A psychological investigation of primigravidae: Part II: Maternal social class, age, physique and intelligence. *Journal of Obstetrics and Gynaecology of the British Empire, LXIII(3)*, 338-343.

Sigel, I. E., & Kelley, T. D. (1988). A cognitive developmental approach to questioning. In J. Dillon (Ed.), *Classroom Questioning and Discussion: A Multi-disciplinary Study.* Norwood, NJ: Ablex.

Sigman, M., & Whaley, S. E. (1998). The role of nutrition in the development of intelligence. In U. Neisser (Ed.), *The Rising Curve*. Washington, DC: American Psychological Association.

Snow, R. E., Kyllonen, P. C., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the Psychology of Human Intelligence, Volume 2* (pp. 47-103). Hillsdale, NJ: Lawrence Erlbaum Associates.

Spicher, S. (1993). *Nouvel Etalonnage du SPM-38: Raven dans le ciel Fribourgeois.* Unpublished manuscript, L'Institute de Psychologie, Universite de Fribourg.

Stallings, J., & Kaskowitz, D. (1974). *Follow Through Classroom Observation Evaluation 1972-1973.* Menlo Park, CA: Stanford Research Institute. Report URU-7370.

Teasdale, T. W., & Owen, D .R. (1989). Continuing secular increases in intelligence and a stable prevalence of high intelligence levels. *Intelligence, 13,* 255-262.

Thorndike, R. L. (1975). *Mr. Binet's Test 70 Years Later.* Presidential Address to the American Educational Research Association.

Thorndike, R. L. (1977). Causation of Binet IQ decrements. *Journal of Educational Measurement, 14,* 197-202.

Tuddenham, R. D., Davis, L., Davison, L., & Schindler, R. (1958). *An experimental group version for school children of the Progressive Matrices.* Unpublished manuscript, University of California. See also Abstract: *Journal of Consultant Psychology, 22,* 30.

United States Government, Bureau of the Census (1984). *Statistical Abstract of the United States, 1983.* Washington, DC: U.S. Government Printing Office.

Vernon, P. E. (1969). *Intelligence and Cultural Environment.* London: Methuen.

Vodegel-Matzen, L. B. L. (1994). Performance on Raven's Progressive Matrices. Ph.D. Thesis, University of Amsterdam.

Vitz, P. C. (1990). The use of stories in moral development. *American Psychologist, 45*(6), 709-719.

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes.* Cambridge, MA: Harvard University Press.

Vygotsky, L. S. (1981). The genesis of higher mental function. In J. V. Wertsch (Ed.), *The Concept of Activity in Society Psychology.* Annank, NH: Sharpe.

Webber, R. J. (1977). *The National Classification of Residential Neighbourhoods. PRAG Technical Paper TP23.* London: Centre for Environmental Studies.

Zhang, H.-C., & Wang, X.-P. (1989). Chinese standardisation of Raven's Standard Progressive Matrices. *Psychological Test Bulletin,* November, *2*(2), 36-39. Hawthorn, Victoria: ACER.

Appendix

Sampling Procedures, Sample Sizes and Data Management

This Appendix outlines some of the considerations which have guided our choice of sampling methodology and data analysis and presentation.

Virtually all statistical tests assume that the groups between which it is desired to discriminate, or from which it is proposed to generalise, are random samples from some wider population. Yet, while attaching much importance to sophisticated statistical technique, psychologists rarely examine the quality of their samples. It is not uncommon for them to assume, for example, that results obtained in studies of psychology students will apply to all people or all people in a category – such as males and females.

Commonly, even when an effort is made to ensure that a population tested is representative of some wider population, "quota sampling" techniques are employed. In these an effort is made to ensure that the demographic characteristics of the population tested correspond to those of some wider population to whom it is hoped to generalise.

Yet, even by the time Hyman wrote his classical book on *Survey Design and Analysis* (1955), it had been repeatedly demonstrated that not only do opinion polls based on huge numbers yield much less accurate data than studies based on much smaller, but randomly selected, samples, so, too, do studies based on quota samples.

For these reasons we have, in our own work, sought to employ systematic random sampling procedures, wherever possible doing so within strata which have been chosen to yield the correct proportions in certain demographic categories required to correspond to wider demographic statistics.

It is important to note that stratification via demographic statistics is a very different matter to asking individual researchers to locate and test specified numbers of people within a number of categories identified in terms of such things as sex, age, socioeconomic status, and ethnic group.

In the 1979 British study we were able, with the aid of funds from the Social Science Research Council, to conduct the study in seven areas of the country which previous research (Webber, 1977) had shown to cover the main variance within the country while at the same time being collectively representative of the country as a whole. We were even able to over-sample particular areas in order to have large enough numbers of respondents to permit detailed comparisons between areas and then re-weight the data to produce have the correct effect when combined with other data in the overall statistics.

In most of the other work summarised in this chapter this has not been possible. It has been necessary to work with collaborators who were interested in contributing to the study and to do what was feasible under the circumstances. As far as possible, we have sought both (1) areas with demographically balanced populations, and (2) a *range* of areas located in parts of the country having very different demographic characteristics. Within areas we have tried to ensure that the samples tested were selected using some strictly random method. In some cases complete lists of names have been obtained and then sampled using a random start and a fixed sampling interval. In other cases it has been necessary to compromise by doing such things as systematically select buildings and classrooms within school districts to be representative of the whole and then test all the children in those classrooms. Such clustering pushes up the numbers but it does not, in fact, yield better samples.

Naturally, data obtained in these ways – unlike those obtained from the 1979 U.K. standardisation among young people – cannot be pooled using routine statistical procedures. Instead, it has to be combined making due allowance for deficiencies in the data set and giving more weight to the more balanced and complete samples.