Cognitive Psychology **41,** 1–48 (2000)
doi:10.1006/cogp.1999.0735, available online at http://www.idealibrary.com on IDEAL®

The Raven's Progressive Matrices: Change and Stability over Culture and Time

John Raven

30 Great King Street, Edinburgh EH3 6QH, Scotland

Data relating to the stability and variation in the norms for the Raven's Progressive Matrices Test (a well-validated measure of basic cognitive functioning) for different cultural, ethnic, and socioeconomic groups on a worldwide and within-country basis are first summarized. Subsequent sections deal with variation over time. A possible explanation for the variation in norms over time and between ethnic groups within countries is offered.

RAVEN'S PROGRESSIVE MATRICES AND MILL HILL VOCABULARY SCALES: THE INSTRUMENTS AND THEIR THEORETICAL SIGNIFICANCE

Raven's Progressive Matrices and Mill Hill Vocabulary Scales were developed for use in fundamental research into the genetic and environmental determinants of ''intelligence.''

Raven (J. C. Raven, 1936; Watt, 1998) set out with the specific intent of developing tests which would be easy to administer and also easy to interpret in a clear, theoretically relevant way. Put another way and with the benefit of hindsight, what he did was make the two main components of ''general intelligence'' (which, as we shall shortly see, have been strongly confirmed in subsequent research) *directly* measurable (as distinct from calculable only by the application of complex, factor-analytically based, weighting procedures) and measurable through using procedures almost as robust and directly interpretable as those used to measure height or temperature.

The tests he produced have been widely applied in both practice and research and a vast pool of data has now accumulated. Inspection of these data appears to reveal that the abilities that are most often thought to lie at the heart of ''intelligence'' are much more open to environmental influence than had previously been thought. However, the factors which influence these

Address correspondence and reprint requests to John Raven, 30 Great King Street, Edinburgh EH3 6QH, Scotland. Fax: ++44 (131) 556 3754.

0010-0285/00 $35.00

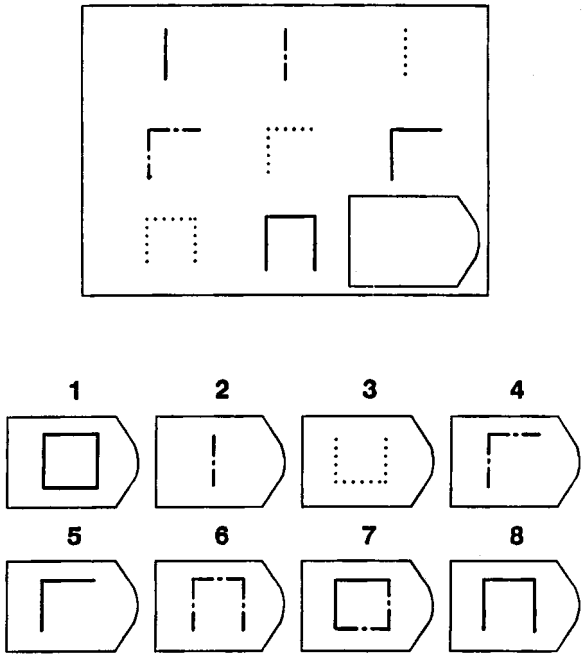

FIG. 1. Illustrative Progressive Matrices item. Respondents are asked to identify the piece required to complete the design from the options below. (The item shown here is not from the current range of tests.)

abilities are *not* those which had previously been expected to be most important by most psychologists and sociologists.

The two main components of general cognitive ability (*g*) which Raven sought to measure directly were those identified by Spearman in 1923 (Spearman, 1927). These are, respectively: (a) *eductive* ability (from the Latin *educere*, meaning ''to draw out''), the ability to make meaning out of confusion, the ability to generate high-level, usually nonverbal, schemata which make it easy to handle complexity; and (b) *reproductive* ability—the ability to absorb, recall, and reproduce information that has been made explicit and communicated from one person to another.

The Raven Progressive Matrices (RPM) tests (of which there are several versions) are made up of a series of diagrams or designs with a part missing. Those taking the tests are expected to select the correct part to complete the designs from a number of options printed beneath. An illustrative item, not from one of the published tests, is shown in Fig. 1.

The basic version of Raven's Mill Hill Vocabulary Scale (MHV) consists of 88 words, arranged in order of ascending difficulty, which these those taking the test are asked to define. The number of words in the test varies with whether those who are expected to take it will be adults or young people.

In the most widely used versions of the test, half the words are in an open-ended format and half in a multiple-choice format.

The theoretical framework which guided the development of the tests has since been confirmed in numerous studies, most recently in work by Horn (1994), Matarazzo (1990), Ree, Earles, and Teachout (1994) and Snow, Kyllonen, and Marshalek (1984). Matarazzo demonstrated that the extraction of more than these two scores from multiple-factor ''intelligence'' tests is usually unjustified. Ree et al. showed that the addition of specific factor scores to *g* estimates rarely improves the ability to predict occupational performance. Snow et al. showed that eductive ability lies at the heart of a Guttman radex, which can be distilled off from many studies and mediates the relationships between verbal, numerical, and spatial reproductive abilities. Horn concluded that: (a) the other eight major factors which have been identified while trying to develop better measures of basic cognitive functioning do *not* merge into either ''fluid'' or ''crystallized'' intelligence before merging to form *g;* (b) that ''crystallized'' intelligence does not ''differentiate out of'' fluid intelligence but is distinct from the start; (c) that these two components of *g,* while correlated, have different genetic origins and are influenced by different aspects of the environment; (d) that they have different developmental trajectories over the life cycle; and (e) that they have different real-life correlates.

It follows from these and other similar results (summarized in, e.g., J. Raven, J. C. Raven, & Court, 1998a, 1998d, 1998e) that: (a) Spearman was probably correct in his formulations and that (b) in reviewing research conducted with the *Raven Progressive Matrices* and *Vocabulary Scales,* one is, in effect, looking at data relating to what must be regarded as the two best established components of general cognitive ability and intelligence more generally.

Not only has the theoretical framework which guided Raven's work been strongly confirmed since the tests were first published, the available evidence also suggests that Raven was successful in developing measures of eductive and reproductive ability. Much of this research has been summarized in J. Raven et al. (1998a, 1998d, 1998e). The factorial evidence is fairly clear, and the Snow et al. review has already been mentioned. Perhaps of more interest here are the findings of Styles and Andrich (1994, 1997) on the one hand and those of a series of researchers—such as Deary (1993, 1995), Deary and Stough (1996), and Vernon (1991, 1993)—who have sought to relate Raven Progressive Matrices scores to ''more basic'' measures of cognitive functioning on the other. By mapping the Item Characteristic Curves (ICCs) for a series of Piagetian tasks onto the set of ICCs for the Standard Progressive Matrices (SPM), Styles and Andrich demonstrated that the development of the ability to give high-level responses to the Piagetian questions is continuous and incremental *and* in step with the development of the ability to solve RPM problems of similar difficulty. Deary (1993, 1995) and others have

shown that RPM scores are linked to measures of inspection time which, it has to be stressed, are, like the RPM, *untimed*, power measures of cognitive functioning.

The overall correlation between the Mill Hill Vocabulary Test and ''intelligence'' tests based on the 'battery of subtests' model (such as the WISC) tends to be in the .8 to .9 range (Court & J. Raven, 1995; Court & C. J. Raven, 1998; Flynn, 1999). The within-age correlation between the Progressive Matrices (RPM) Tests (and, indeed other measures of eductive ability) and the Mill Hill Vocabulary (MHV) Scale (and other measures of reproductive ability) tends to be of the order of .5 (Court & J. Raven, 1995).

Versions of the Tests

Most, but not all, of the research to be summarized in this article was conducted with the Standard (as distinct from the Coloured or Advanced) Progressive Matrices Test. It is important to note that the SPM was, from the start, known to have both certain strengths and limitations. Its strengths were that it could be used with respondents of all ages from early childhood to old age and was of such a length that it could reasonably be administered in homes, schools, and workplaces (where time is necessarily limited) as well as in laboratories. It was thus particularly useful for comparative studies. Naturally, it had limited discrimination at the upper and lower levels. This was overcome by developing the Advanced (APM) and Coloured (CPM) Progressive Matrices Tests for use among the more and less able, respectively.

Creeping Awareness of the Importance of Studying Change Over Time

One of the main aims of this article is, following the attention drawn to the phenomena by such authors as Thorndike (1975, 1977), Schaie (1983), Schaie and Willis (1986), and Flynn (1984, 1987), to review the evidence available from researchers who have worked with the Raven Progressive Matrices and Vocabulary Scales which bears on the still-controversial question of whether there has been a major increase in ''intelligence'' over the past century. However, since it will emerge that, as is also evident from Flynn's writings from 1984 to 1999, the answer to this question depends centrally on what is *meant* by ''intelligence,'' it is necessary to examine the results obtained with the two tests separately and to ask whether any trends found are universal or confined to certain ability, cultural, or ethnic groups.

Given that the tests have been in use for more than 60 years, distilling off the evidence bearing on the questions just raised is not so easy as might be expected. Since most psychologists never even suspected the effect, they not only did not think it was necessary to collect data which bore on the hypothesis, they simply assumed that normative data collected in the past were still applicable. They did not see any need to restandardize tests. Another factor is that in cross-sectional studies evidence for changes in a trait

across birth cohorts is inevitably correlated with changes in the trait across age. Although some researchers (e.g., Owens, 1966; Thorndike, 1975) using other tests noted an apparent increase in the scores on some components of ''intelligence,'' they failed to note that the eductive component was increasing at a dramatic rate. Garfinkel and Thorndike (1976) and Schaie and his colleagues (Schaie, 1983, 1994; Schaie & Strother, 1968) deserve credit for observing cohort-related changes in different components of intelligence. However, this work did not make a great impact outside of gerontology. Flynn (1984, 1987, 1989) deserves the credit for the systematic analysis of cohort effects as revealed by data on the use of the RPM as a screening device for military inductees in a number of countries.

There are substantial methodological problems associated with this effort. First, to be meaningful, the data had to be sectioned by age, as in Table 1 (which will be explained more fully later). Second, the bimodal and skewed within-age distributions shown in Fig. 2 (redrawn from J. Raven, 1981), combined with a scatter which varied with age (also illustrated in Fig. 2), meant that the normal data-reduction techniques could not easily be applied.

I now consider cohort changes as revealed by studies of the standardization samples for the various Raven Progressive Matrices and Vocabulary tests.

STANDARDIZATIONS AMONG YOUNG PEOPLE

The Standard Progressive Matrices (SPM) was first fully standardized by J. C. Raven on 1,407 children in Ipswich, England, in 1938 (J. C. Raven, 1941).[1] 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 out in a town not far from Ipswich, namely, Colchester, over the years 1943–1944. Perhaps because of the disruption caused by World War II, the SPM norms obtained in that study were consistently 2 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 1963–1965 Skanes tested 4,017 children ages 9 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 ages 10 to 14 years. The results consistently lagged *behind* the Ipswich norms. In 1972, Byrt and Gill (1973), working with the

[1] Compilation of these norms is continuing. The author would welcome contact from anyone interested in contributing to the process.

TABLE 1
Standard Progressive Matrices: 1979 British Percentile Norms for the Self-Administered or Group Test among Young People (Smoothed)

	Age in years (months)																		
Percentile	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)	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)
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.

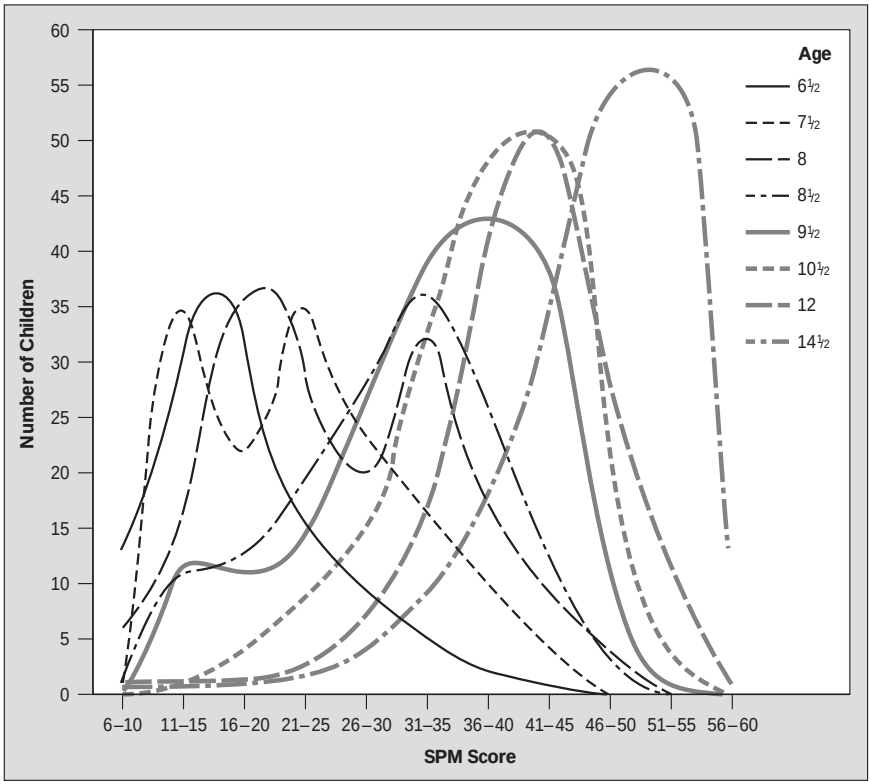

FIG. 2. Distributions of raw scores for eight age groups on the Standard Progressive Matrices. Data from the 1979 British Standardization among young people.

author, collected data from a nationally representative sample of 3,464 primary school children ages 5 to 11 years in the Republic of Ireland. The urban norms seemed to corresponded to the 1938 Ipswich norms, although the figures for the 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 onward 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 UK 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 for what was then Czechoslovakia which corresponded to a recently reported British study. Work carried out in the United States by J. Raven et al. (1990/2000) revealed that, while the overall U.S. norms lagged behind these 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 those obtained elsewhere. Still more recently, similar data have been reported for Poland (Jaworowska & Szustrowa, 1991; J. Raven, J. C. Raven & Court, 1998c, 1998d), Spain (J. C. Raven, Court, & J. Raven, 1995), further school districts of the United States (J. Raven & Court, 1989), and Switzerland (Martinolli, 1990; Spicher, 1993).[2]

Two observations may be interjected at this point. First, when reporting the results of the 1979 British standardization, 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 just been summarized, 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.

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 now summarize 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 are reviewed first, followed by a review of studies of variance within countries.

As has been mentioned, Chan's Hong Kong norms exceed most of the

[2] Although the RPM data are limited to 60 years, Tuddenham's (1948) army data go back to 1914 and the Binet data go back further still (Thorndike, 1975).

norms already discussed. However the norms which most significantly exceed them come from Taiwan (Miao & Huang, 1990; Miao, 1993). A possible explanation of these results has been presented elsewhere (Raven et al., 1990/2000).

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 (Sahan & 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 (J. Raven et al., 1998b).

As emphasized 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. standardizations 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 STANDARDIZATION

The 1979 British Standardization was conducted in seven areas of the country which were chosen, under the guidance of the Government Office of Population Censuses and Surveys, to represent all the types of area into which a cluster analysis of large amounts of demographic data had shown the variance within the country could be classified (Webber, 1977). The types of area in which few people lived were oversampled in order to have enough respondents to make it possible to break the data down by type of region. Later, the data were reweighted to its correct proportions to give overall statistics. It was therefore possible to employ fairly sophisticated statistical procedures when analyzing the data. Altogether 3250 children ages 6 to 16 were tested.

Previous research had shown that eductive ability tends to be related to socioeconomic status (SES). The questions used to elicit the information used to classify SES, the rationale for employing these questions, and the framework used to classify the results are discussed in Appendix A.

This is a convenient point at which to explain the format in which the data will be displayed. Many authors present relevant data in terms of Deviation IQs with a mean of 100 and a standard deviation of 15. This process

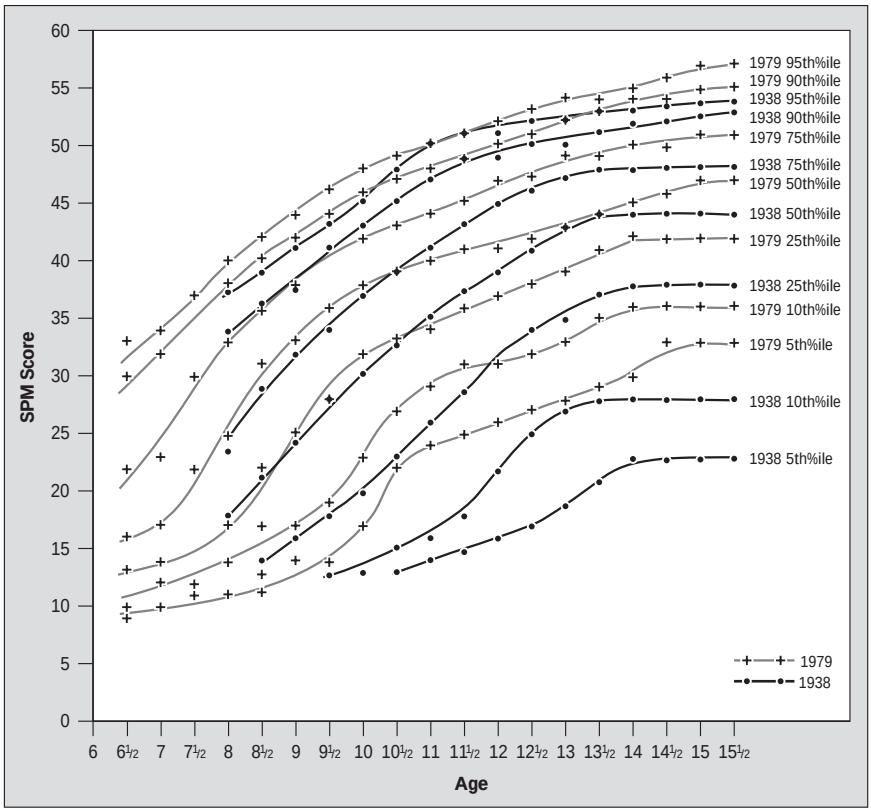

FIG. 3. Graphed percentile norms for young people in Great Britain on the Standard Progressive Matrices in 1938 and 1979. The graphs show the scores obtained by young people of different ages and levels of ability in these 2 years. If one compares the graphs of the 1938 norms (i.e., the heavy lines) 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, 10½-year-olds in 1979 obtained similar scores to those obtained by 14-year-olds in 1938.)

is, in general, unjustifiable for a number of reasons which include two that are important here: First, as was evident from Fig. 2 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.

Table 1 presents the raw scores corresponding to the 5th, 10th, 25th, 50th, 75th, 90th, and 95th percentiles, by age, in the 1979 British standardization group. The data are also presented in Fig. 3. Note that (as discussed more fully in Appendix B) these scores can be quite unreliable at the lowest and

TABLE 2
Standard Progressive Matrices: Contribution of SES, Sex, and Age to Total Variance (1979 British Data for Young People)[a]

	Simple *R*	Mult. *R*	R^2	R^2 change	β
Age	68	68	46	46	67
Sex	1	68	46	0	0
Father's SES	22	71	50	4	17
Region	16	71	50	0	7
Age	68	68	46	46	67
Sex	1	68	46	0	0
Region	16	69	48	1	7
Father's SES	22	71	50	2	17

Note. $n = 3256$.
[a] Decimal points omitted and rounded to two decimal places.

highest percentiles, even though the groups may be of reasonable size. For example, the 95th percentile of a group containing 50 people is determined by the midpoint of the scores obtained by the second and third highest scorers. For this reason, the presentation of unsmoothed raw data can lead to a quest for explanations of chance fluctuations and, when used as reference data against which to view the scores of individuals or experimental groups, to seriously misleading evaluations. Unless otherwise stated, all data presented in this article have therefore been smoothed by graphing.

As can be seen from Table 2, 2.6% of the SPM variance was accounted for by region, but when the effect of SES was partialed out, this dropped to .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 SES 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 generalize from one population to another.

The 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 standardization, 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 consume too much space to present here, it may be noted that: (a) the ICCs for individual items were remarkably similar to those published 40 years earlier and that (b) as can be seen from the summary data presented in Table 3, the items scaled in much the same way for children from a variety of different backgrounds.

TABLE 3

Standard Progressive Matrices: Correlations between Item Difficulties Calculated Separately for Young People from Different Socioeconomic Backgrounds (Data from 1979 British Standardization)[a]

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	

[a] Decimal points omitted and rounded to two decimal places.

The conclusion is clear and vitally important: It is not possible to explain away differences between 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 are not 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. STANDARDIZATIONS AMONG YOUNG PEOPLE

Between 1983 and 1989 some 50 norming studies were carried out within school districts spread across the United States (J. Raven et al., 1990/2000; J. Raven, 1989). 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 ages 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 4 and 5 both ethnicity and socioeconomic status seemed to make independent contributions to the within-district variances. It is, of course, not possible to establish this point with absolute certainty since SES is correlated with ethnicity. All we were able to do was, as shown in Table 5, to run the regressions twice, once with SES partialed out first and once with ethnicity partialed out first. When this was done, whichever way the variables were entered, there was variance left to be explained by other variables.

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 6. 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 Fig. 4, Hoffman (1983, 1990) demonstrated that the regression lines of RPM on various types of *achievement* for different ethnic groups were (to all intents and purposes) parallel—although having different intercepts. (Although the regression lines for mathematics shown in Fig. 4 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 reported by Jensen (1974).

THE 1979 STANDARDIZATION AMONG YOUNG PEOPLE IN GREAT BRITAIN

The Mill Hill Vocabulary Scale

The Mill Hill Vocabulary Scale (MHV) was standardized alongside the SPM in the 1979 study which has already been described. As can be seen from Table 7, and as was also the case with the SPM, there was no variance in MHV scores with region once the effect of SES was partialed out.

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 the same way as SPM scores; growth continued at approximately one and a half words per 6-month interval through to age 15 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.

U.S. Data for Young People

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, 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

TABLE 4
Standard Progressive Matrices: 1986 Adolescent Percentile Norms for Ethnic Groups in ''Westown'' (United States) in the Context of 1979 British Data (Smoothed)

Percentile		Age in years								
		12½	13	13½	14	14½	15	15½	16	16½
	UK	53	54	54	55	56	57	57	—	—
	Anglo	51	52	53	54	55	56	57	58	59
95	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
	UK	51	52	53	54	54	55	55	—	—
	Anglo	50	51	52	53	54	55	56	57	57
90	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
	UK	47	49	49	50	50	51	51	—	—
	Anglo	46	47	47	48	50	52	53	54	54
75	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
	UK	42	43	44	45	46	47	47	—	—
	Anglo	41	42	43	44	47	48	48	48	49
50	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
	UK	38	39	41	42	42	42	42	—	—
	Anglo	37	38	39	40	42	44	45	45	45
25	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
	UK	32	33	35	36	36	36	36	—	—
	Anglo	32	33	34	35	36	38	40	40	40
10	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
	UK	27	28	29	30	33	33	33	—	—
	Anglo	27	28	29	30	32	34	36	37	38
5	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
	UK	174	185	180	196	189	191	171	—	—
n	Anglo	46	59	44	52	53	36	56	40	49
(unweighted)	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.

TABLE 5
Standard Progressive Matrices: Contributions of Ethnicity and SES to Total Variance 1986 Data for Adolescents in ''Westown'' (United States)[a]

	Simple *R*	Mult. *R*	R^2	R^2 change	β	β^2
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
Asian	−04	44	19	0	0	0
Father's SES	−31	48	23	4	−20	4

Note. Ethnic groups were entered as dummy variables, i.e., coded as ''yes'' or ''no.''
[a] Based on weighted data; decimal points omitted and rounded to two decimal places.

TABLE 6
Standard Progressive Matrices: Correlations Between Item Difficulties Calculated Separately within Specified Groups[a]

	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

[a] Decimal point omitted and rounded to two decimal places.

other cultures. As with the SPM, and as can be seen from Table 9, 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 words included in the Scale that are unknown to other cultural groups.

ADULT STANDARDIZATIONS

Standardization in the United Kingdom in the Mid-1940s

The UK adult norms for the SPM—which formed the main reference data used worldwide for more than half a century—were built up from a number of sources.

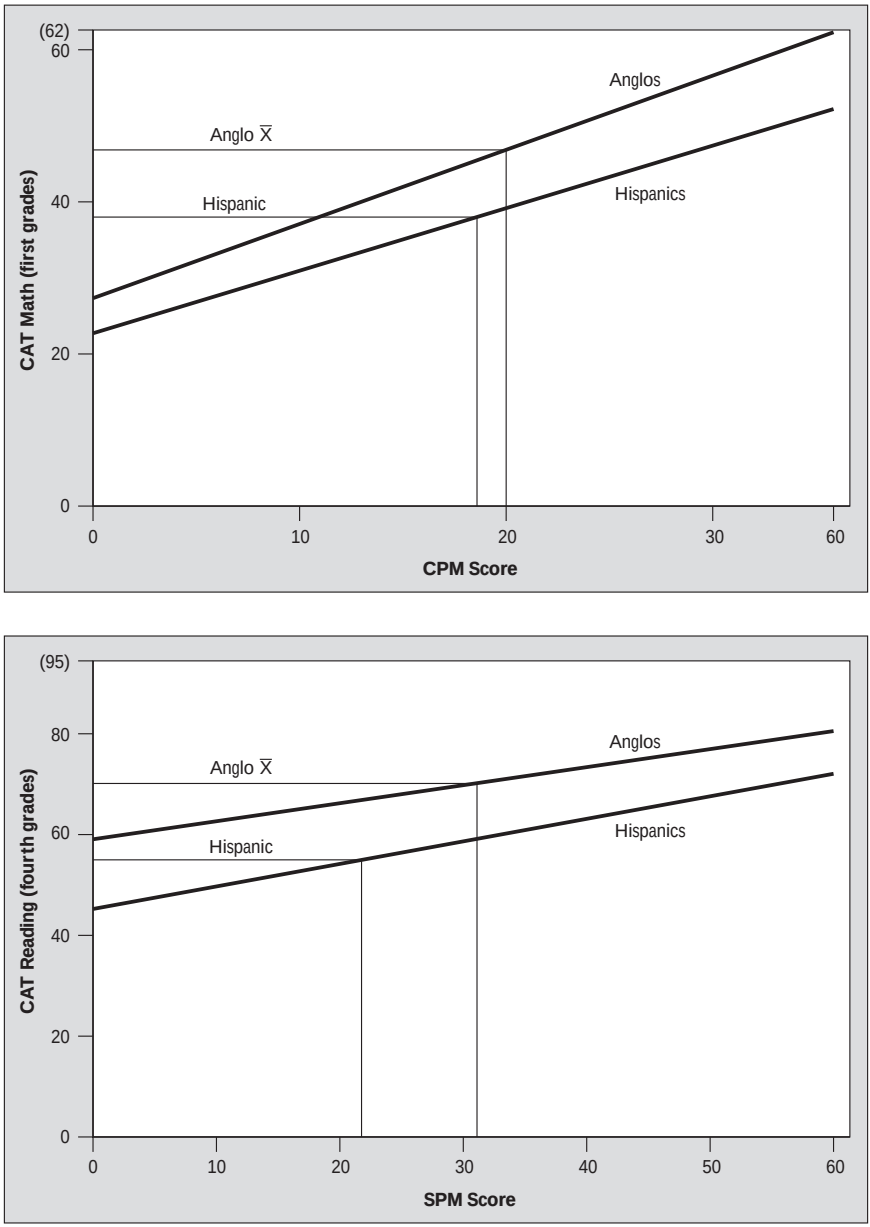

FIG. 4. Sample regressions of the subtests of California Achievement Test on the RPM for Anglos and Hispanics in Douglas, Arizona. The upper figure shows the regressions of CAT Math scores on the Coloured Progressive Matrices among first-grade students. The lower figure shows the regressions of CAT Reading scores on the Standard Progressive Matrices among fourth-grade students. (Redrawn from Hoffman, 1990.)

TABLE 7
Mill Hill Vocabulary Scale: Contribution of SES, Sex, and Age to Total Variance (1979 British Data for Young People)[a]

	Simple *R*	Mult. *R*	R^2	R^2 change	β
Age	76	76	58	58	74
Sex	0	76	58	0	−01
Father's SES	26	79	62	5	20
Region	17	79	63	0	7
Age	76	76	58	58	74
Sex	0	76	58	0	−01
Region	17	77	59	1	7
Father's SES	26	79	63	4	20

Note. $n = 3382$.
[a] Decimal points omitted and rounded to two decimal places.

As described by J. C. Raven (1941), the first group consisted of the fathers of the children tested in the course of the Ipswich study mentioned above. The second group consisted of conscripts and other recruits to the World War II British armed services. All these men were physically fit, and those from civilian occupations which were ''reserved'' and others with congenital disabilities were excluded. All were aged 20–30, the majority being under 25. Since the SES distributions of the Ipswich fathers and armed services personnel were similar, the two groups were combined and the total of 3665 was treated as a representative cross section of adults of that age range. There were marked differences in the scores obtained by men having different kinds of occupational and educational backgrounds.

The next study, conducted in 1946 (Foulds & J. C. Raven, 1948), was carried out to examine trends in the scores with advancing age [which could not be achieved with the (young) armed service personnel]. All employees of the Post Office engineering division (Scotland) and a London manufacturing company were asked to take part in the study. About half the Post Office employees agreed to do so (although the proportion fell to about a third among those ages 50 and over). About a third of the employees of the manufacturing company agreed to take part. Altogether, 1967 men participated in the study. Graphs like those shown in Fig. 5 were plotted separately for the percentile age norms obtained from the participants from the two employers.

When the occupational groups and SPM scores of these volunteers were compared with those of the 25-year-olds in the combined Ipswich and military group, they were found to be considerably higher. More specifically, few of those employed by these organizations obtained scores corresponding to those obtained by the bottom 10% of the Ipswich-military group. Detailed examination of the data made it clear that the age trend established for the bottom 5th percentile of the civilian group corresponded to that which would

TABLE 8
Mill Hill Vocabulary Scale: 1986 Adolescent Percentile Norms for United States 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. U.S. figures estimated on the basis of data available summer, 1986. The studies on which these norms are based are detailed in Raven et al. (1990/2000). These show that the norms vary considerably between school districts, and, within districts, between ethnic groups.

TABLE 9
Mill Hill Vocabulary Scale: Correlations between Item Difficulties Calculated Separately within Specified Groups[a]

	Westown Black	Westown White	Westown Hispanic	Westown Asian
Westown Black				
Westown White	97			
Westown Hispanic	100	97		
Westown Asian	90	97	99	

[a] Decimal point omitted and rounded to two decimal places.

have been obtained for the 10th percentile of the general population. No age trends for the general population 5th percentile could therefore be established. The general trends with age for the differing percentiles established in the civilian data were then re-plotted, starting the graphs at the appropriate score derived from the Ipswich-military data.

As indicated in J. C. Raven (1948), a further 1145 male postal workers were added the following year to these standardization groups to yield the adult SPM norm table included in all editions of the guides to, and subsequent manuals for, the SPM until 1992. Figure 5 shows these figures in the form of a graph. Although, as has been indicated, the data were collected over a number of years, they will, for convenience, hereafter be referred to as the ''1942 UK adult norms.''

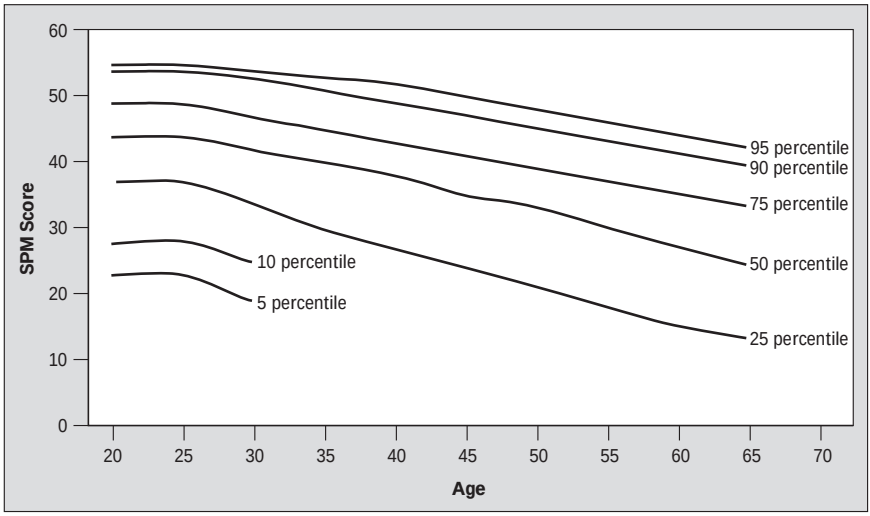

FIG. 5. A typical figure showing the apparent decline in Standard Progressive Matrices Scores with increasing age among people of different levels of ability as obtained in cross-sectional studies. As described in the text, the data were accumulated in the course of a series of studies conducted between 1939 and 1947.

1992 STANDARDIZATION IN DUMFRIES, SCOTLAND

In 1992 a further standardization of the RPM was conducted in Dumfries, Scotland. The study made use of several previously established facts: (a) the studies among schoolchildren discussed earlier had made it clear that the most important variable to take into account was the balance of SES groups in the population studied; (b) the 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 United Kingdom as a whole—*did* correspond to those of the UK as a whole; and (c) the town of Dumfries recommended itself as a possible site for an adult standardization because (1) much of the data collected with the RPM over the past 50 years (including two major standardizations of the Coloured Progressive Matrices) had been collected there and had held up well in international comparisons and over time, (2) it had itself a demographic structure which approximated that of the United Kingdom, and (3) 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 as well as response rates are fully described in J. Raven et al. (1998c, 1998d). Here it is sufficient to note:

1. That because of the ceiling on the test which had become notable in the previously mentioned studies of young people, it was necessary to standardize the APM alongside the SPM. To do this, APM Set I was administered by the researcher on contacting the named respondent and scored immediately. Depending on that score *either* the SPM or APM Set II was left for completion when the respondent had time. Test booklets and answer sheets were later collected by the researcher. If the test completed was the APM, the scores were converted to SPM equivalents using a table established by Andrich and Dawes (now Styles) (1989). The effect of leaving the tests for completion at leisure was later shown to be minimal (J. Raven, 1995; J. Raven et al., 1998d) (A summary of this evidence will be found in Appendix C).
2. Stratified systematic random, not quota, sampling procedures were used because, despite their greater cost, they have been shown to yield much more accurate results (Hyman, 1955; see also Appendix B).

Just over 20% of those contacted declined to take part in the study. As in the earlier work, this proportion increased among the more elderly. Altogether 645 respondents were tested.

1993 STANDARDIZATION IN DES MOINES, IOWA

Following the Dumfries study, Chaplik, Berrill and others (See J. Raven et al., 1998c, 1998d) conducted an exactly parallel study in Des Moines,

Iowa. Des Moines is recognized as one of four U.S. cities having demographic compositions approximating the United States as a whole and is therefore widely used by researchers seeking a microcosm of America (*American Demographics*, May, 1985). While it is, of course, impossible for any one city—however closely its crude demographic statistics may correspond to the whole country—to correspond to the whole at a detailed level, the studies conducted with the RPM among schoolchildren in the United States in the 1980s had in fact confirmed that the norms for Des Moines did approximate to those for the United States as a whole (J. Raven et al., 1990/2000).

There were two major differences between the Dumfries and Des Moines studies. First, Des Moines is much bigger than Dumfries. It was therefore necessary to sample areas within Des Moines before proceeding to identify individuals to be contacted. Judy Connor and Gary White of the sampling section of the Institute of Social Research in Michigan used their database to select 30 areas of Des Moines which, collectively, yielded a picture (in demographic terms) of Des Moines as a whole as well as the variations between different types of areas within it. Second, there is, in the United States, no list of residents as complete as the electoral registers in the United Kingdom. It was therefore necessary to follow procedures developed by Kish (1949) to identify the individuals to be tested.

Seventy percent of those contacted agreed to take part. Six hundred twenty-five complete sets of data were obtained. Twelve percent did not give sufficient information to enable their socioeconomic status to be ascertained. Of the remainder, 23% had professional and managerial occupations. This compares with 21% of Des Moines residents, and 20% of U.S. residents, as reported in the 1983 *County and City Data Book*. Only 4% did not give sufficient information on their ethnicity to make classification possible. Of those who did provide this information, 8% indicated they were Black, 87% White, and most of the rest Hispanic or Asian. The 1990 census figures for Des Moines are 7% Black and 89% White and, for the United States as a whole, 12% Black and 83% White. These figures suggest that there is every reason to believe that the sample tested is indeed representative of Des Moines and that, in terms of relevant variables, its demographics approximate those of the United States as a whole.

COMPARISON OF DUMFRIES AND DES MOINES DATA

The APM norms for Des Moines are compared with those from Dumfries in Table 10. 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 United Kingdom, while the 50th and lower percentiles—and especially the latter—lag behind, at least up to age 50. These data

TABLE 10
Advanced Progressive Matrices, Set II (Untimed): 1993 Adult Percentile Norms for Des Moines, Iowa (United States) in the Context of 1992 Dumfries (United Kingdom) Data

	Age in years																					
	20 (18 to 22)		25 (23 to 27)		30 (28 to 32)		35 (33 to 37)		40 (38 to 42)		45 (43 to 47)		50 (48 to 52)		55 (53 to 57)		60 (58 to 62)		65 (63 to 67)		70 (68+)	
Percentile	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.

suggest that one can have considerable confidence in the meaningfulness of the test scores and the sampling and testing procedures used in both Dumfries and Des Moines. Nevertheless, our extensive research among young people (J. Raven et al., 1990/2000; J. Raven & Court, 1989) does suggest that the lower percentiles for the United States as a whole should lag further behind the UK norms than those obtained in Des Moines. A number of possible explanations of this are discussed in J. Raven et al. (1998c) but, whatever the explanation, the 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 11 shows that the adult norms for the MHV for Dumfries and Des Moines were also similar.

STABILITY AND CHANGE OVER TIME

Our next step is to review the evidence of change and stability in scores with date of birth in more detail.

Figure 3 displays the 1979 normative data for the SPM which were derived from the previously described nationwide study of young people in Great Britain (presented in Table 1) in the context of the data obtained in the 1938 Ipswich study. It is important to emphasize that these graphs 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. It would appear that, by 1979, young people mastered the abilities tested by the Matrices at an earlier age, and that the scores—particularly of the less able—plateaued above their previous level. Not only had there been an increase in the level at which the scores plateaued, children attained given levels of score much earlier.

More recently Martinolli (1990) with the CPM and Spicher (1993) with the SPM have demonstrated similar changes in the norms over time in Fribourg, Switzerland.

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 within the past decade.

Bouvier's (1969) data, also based on military recruits and reproduced in Fig. 6, suggest that the increase had been steady rather than associated with particular developments.

The SPM results from the 1992 adult standardization in Dumfries are shown, plotted by date of birth, by the dotted lines in Fig. 7. The dashed lines replot the scores obtained in the 1940s study previously shown in Fig. 5.

Examination of the points at which the two sets of graphs interface (i.e., for people born in 1922, where the earlier data are particularly strong) reveals that both the mean and the spread of scores were very similar regardless of

TABLE 11
Mill Hill Vocabulary Scale, Forms 1 and 2 (Untimed): 1993 Adult Percentile Norms for Des Moines, Iowa (United States) in the Context of 1992 Dumfries (United Kingdom) Data

	Age in years																					
	20 (18 to 22)		25 (23 to 27)		30 (28 to 32)		35 (33 to 37)		40 (38 to 42)		45 (43 to 47)		50 (48 to 52)		55 (53 to 57)		60 (58 to 62)		65 (63 to 67)		70 (68+)	
Percentile	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.

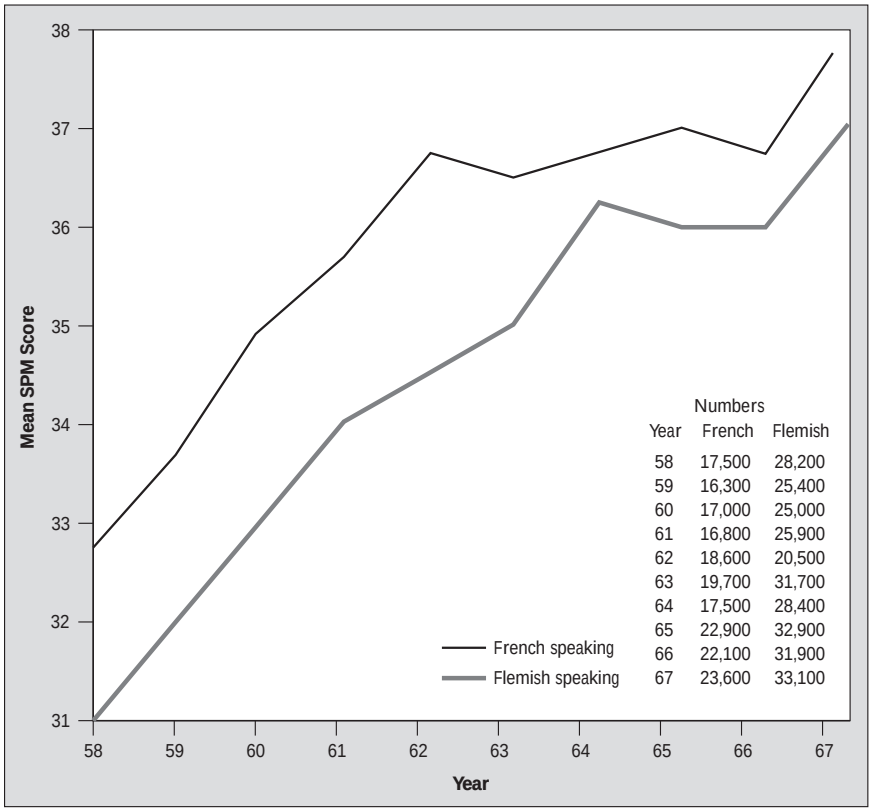

FIG. 6. Mean scores of Belgian military conscripts on the Standard Progressive Matrices from 1958 to 1967. The mean scores of French- and Flemish-speaking recruits are graphed separately. (Redrawn from Bouvier, 1969, and reprinted with permission.)

whether they were derived from the sample tested in the mid to late 1940s (when they were 20 years old) or from the sample tested in Dumfries in 1992 (when they were 70 years old). Instead of a decline in scores with age, 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 Fig. 7 in Fig. 8. The graphs for the Heron and Chown data run straight through the point of interface between the graphs for

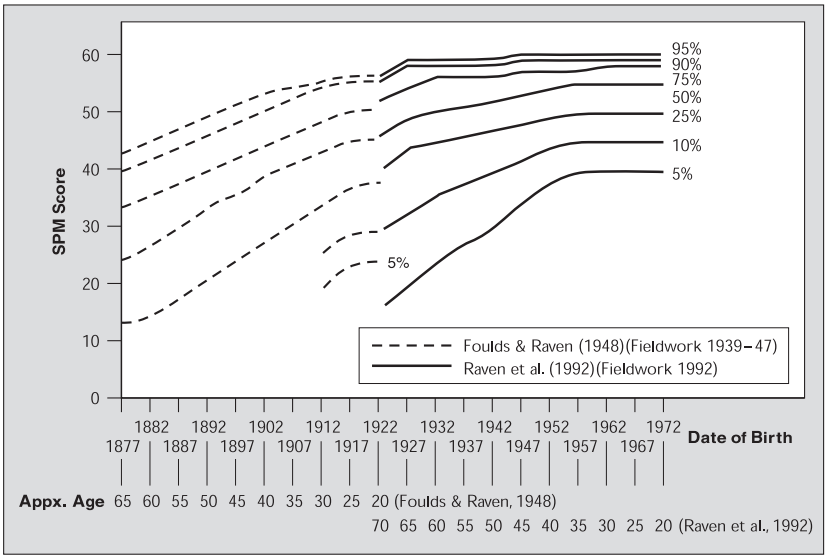

FIG. 7. One hundred years of eductive ability. The figure graphs the percentile norms obtained by adults of different ages (and thus dates of birth) on the Standard Progressive Matrices when a sample was tested circa 1942 (see legend to Fig. 5) in one case and in 1992 in the other. The approximate age of people born in different years in the two samples is shown below. It will be seen that those born in 1922 and tested circa 1942 (and thus approximately 20 years of age when tested) obtained similar scores to those born in 1922 and tested in 1992 when they were 70 years of age.

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 Figs. 7 and 8, and because these figures apparently confirm Teasdale and Owen's (1989) statement that ''. . . we 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 standardizations, the Advanced Progressive Matrices was standardized alongside the SPM. The APM norms for Dumfries for 1992 are compared with the 1962 norms for the same test in Table 12. It is immediately obvious that the increase in scores evident in the lower percentiles in Fig. 7 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 in the scale have been discussed by Prieler, 1998) 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

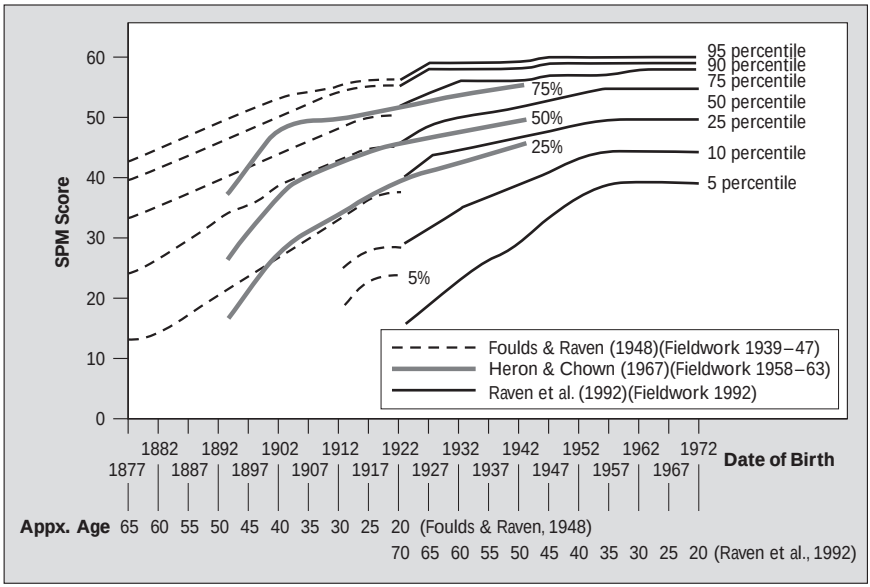

FIG. 8. One hundred years of eductive ability, including data from 1962. Data from a further study conducted approximately halfway between the other two have been superimposed on those displayed in Fig. 7. (For a fuller explanation see legend to Fig. 7.)

TABLE 12
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.

ability have moved up (with admittedly some change in shape), so has the entire distribution of eductive ability.

Turning now to *reproductive* ability, Table 13 compares the 1979 UK norms for adolescents on the MHV in the United Kingdom with those obtained using the written test in Colchester in 1943. The 95th percentile has

TABLE 13
Mill Hill Vocabulary Scale, Forms I and II (Self-Completed in Writing): 1979 British Adolescent Norms 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 1419 (1943 data) and 1304 (1979 data).

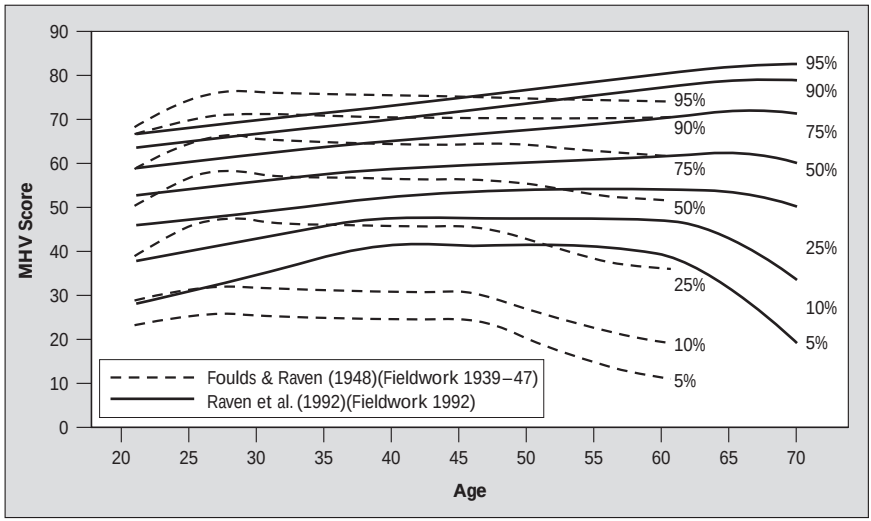

FIG. 9. Stability and change in adult reproductive ability. The graphs plot the percentile Mill Hill Vocabulary Scale scores achieved by a cross section of adults tested in the mid-1940s in the context of those achieved in 1992. It will be seen that, although there appear to have been some changes—with the scores obtained by less able middle-aged adults seeming to have gone up most—the changes are not as great as those which have occurred in eductive ability.

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

Figure 9 presents the UK adult data. It suggests that there has been very little change in the MHV scores obtained by average and above-average adults over the half-century between the two large-scale standardizations reported above.

Bouvier's (1969) data (Fig. 10) likewise reveal little change in vocabulary test scores over the period of his study—and especially among the French-speaking group.

All these results suggest that reproductive ability—and, more specifically, 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.

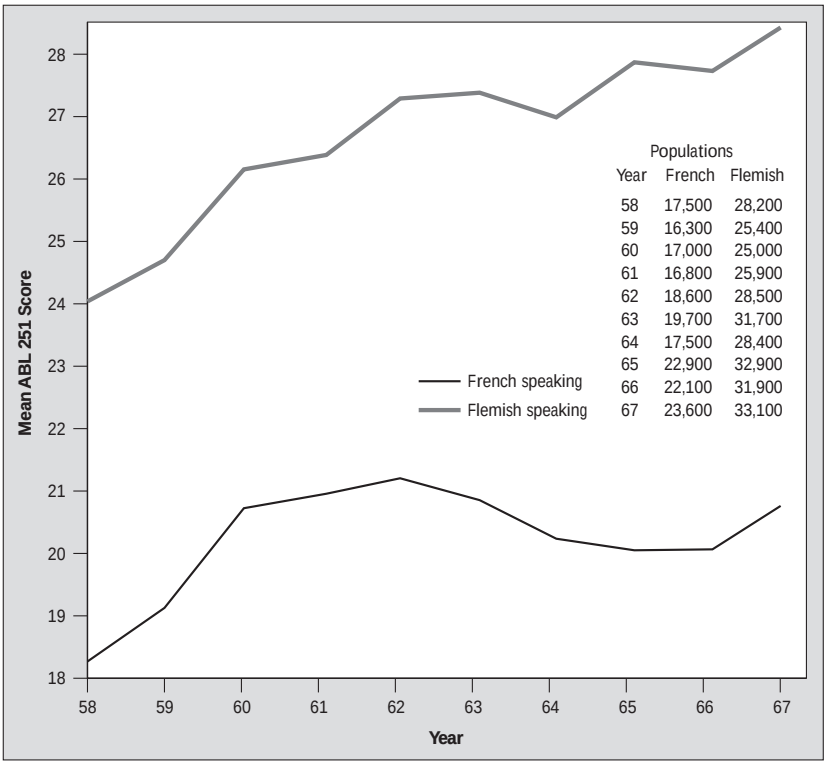

FIG. 10. Mean scores of Belgian military conscripts on the Belgian Army Vocabulary Tests (ABL 251) from 1958 to 1967. The mean scores of French- and Flemish-speaking recruits are graphed separately. (Redrawn from Bouvier, 1969, and reprinted with permission.)

Schaie (1983, 1994) and Thorndike (1977) have 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 extents have increased in proportion to the extent to which they involve eductive ability. Schaie's position on these issues seems to have shifted over the years. In earlier publications, he (Schaie, 1983; Schaie & Willis, 1986) 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 gradi-

ents'' 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 summarized 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 & 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 summarizing 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 are 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 a possible explanation of the changes over time make that explanation 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. (1990/2000) 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 differ-

ence between the RPM norms obtained in Scotland and England in the 1979 standardization 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 8 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 1 point of the 20-point IQ 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 helps 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 behooves 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 United States does seem to correspond to variations 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 (Knight & Eldridge, 1984; 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. Such evidence as we were able to garner (summarized in J. Raven

et al., 1998a) did seem to support this hypothesis but, since the outcome was far from certain, it has been eliminated here.

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 hereditarian 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.

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 realize 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 empathize 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 Oates (1999), 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 intergenerational 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.

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 Fig. 7 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 organizations 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.

APPENDIX A

Classification of Socioeconomic Status (SES)

The classification of socioeconomic status is always difficult, and the more so when it has to be based on data obtained from young children. Children rarely know their parents' incomes (never mind whether before or after tax) and further details on household socioeconomic status, routinely collected in surveys, are both hard to come by and experienced by children as intrusive and irrelevant to the purpose of the study. Data on mothers' occupations are always hard to interpret, hard to combine with data from their partners, and tend, in any case, to cluster into secretarial and service occupations.

It may be thought that these difficulties would invalidate any attempt to use SES as an analytic variable in social research. Surprisingly, this is far from being the case. While some 25% of children report that they do not know their father, have an unemployed father, do not know his job, or do not provide sufficient information to permit classification, the information on fathers' occupation returned by the remainder has proved persistently—indeed disconcertingly—useful. It not only predicts school success better than full-length intelligence tests, it predicts life success very much better. There have been many reluctant converts to the use of a simple index of SES in research. These include both the author and J. Newson and E. Newson (1965). The latter embarked on their research with a commitment to breaking

with tradition and assessing separately (and determining the differential impact of) home process variables which contribute to the general factor of SES that is generally indexed by father's occupation. However, they found themselves forced to conclude not only that a simple index of occupational prestige had as much explanatory power as all the other information so cumbersomely obtained and combined, but also that one learned little from the attempt to study the relative contribution of each identifiable variable.

This is not altogether surprising. Kohn, Slomczynsky, and Schoenbach (1986), following in the footsteps of such researchers as Warner and Lunt (1941), Warner, Meeker, and Eells (1949), Hall and Jones (1950), Inkeles and Rossi (1956), and Berelson and Steiner (1964), demonstrated that a simple index of fathers' occupational status is the best single measure of a general factor of family socioeconomic status based on four indices of fathers' occupational status (Duncan, 1961; Hollingshead, 1967; Hodge, Siegel, & Rossi, 1964; Treiman, 1977), mothers' occupational status, incomes of both parents, and parents' levels of education. Fathers' occupational status had a factor loading of .86. Significantly, too, they also demonstrated that the factor structure and its indices are similar in Poland, the United Kingdom, and the United States.

While, therefore, much remains to be done by way of explaining the relationship between SES and a host of psychological and sociological variables, it is not only impractical to base the classification of SES on more information than we would have been able to collect in the studies summarized in this article, collection and composition of such information would not be expected to yield a ''better'' overall index. Furthermore, analysis of the relative contributions to RPM variance that are made by the component variables treated separately would not, on the basis of past experience, be expected to greatly add to the insights achieved.

Extensive pilot work conducted by the author while working at the UK Government Office of Population Censuses and Surveys revealed (Atkinson, 1968) that two separate questions are required to elicit the necessary information from young people: (1) ''What is the name of your father's job?'' and (2) ''What exactly does he do in that job?'' Illustrative answers would be ''Civil Servant'' and ''He sweeps the floors.'' These questions were employed when collecting background data for the young people involved in most of the studies that were conducted by, or in collaboration with, the author that have been summarized in this article. The answers were classified into the Hall–Jones framework. (For further information see J. Raven, 1981.)

APPENDIX B

Sampling Procedures, Sample Sizes, and Data Management

This Appendix outlines the considerations which 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 generalize 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.

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 generalize.

Yet, even by the time Hyman wrote his classic 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 oversample particular areas in order to have large enough numbers of respondents to permit detailed comparisons between areas and then reweight the data to produce the correct effect when combined with other data in the overall statistics.

In most of the other work summarized in this article 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 then 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 UK standardization 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.

There is one more matter which merits comment. As Deming (1980) has been at pains to observe, human beings without a firm background in science are wont to seek explanations for variation, however meaningless an examination of the nature of the overall situation shows that variation to be. In our own work this sometimes expresses itself in a quest for the ''raw data'' despite the fact that those data obscure what is really going on. Just how great such irrelevant variation can be may be seen from a glance at Figs. 6 and 10. These graphs are based on huge numbers which are not correctly described as samples but as studies of virtually everyone in the populations concerned. It follows that the ''noise'' variance does not arise from sampling and would, given the huge numbers, be overwhelmingly ''statistically significant.'' When the data are based on smaller numbers and sampling error comes into play, the results are, as can be seen from Table B1, even more irregular. (The extent of the problem is immediately apparent if one traces the raw scores corresponding to the 5th percentile across the page. Note, too, that the figures are based on an unusually large and carefully drawn sample.) The solution to this problem is to smooth the data by the time-honored scientific process of graphing. All data presented in the main tables in this article have therefore been smoothed in this way.

APPENDIX C

The Effect of Leaving the Tests for Completion-at-Leisure in the Adult Studies in Dumfries and Des Moines

Questions are frequently asked about the effect of leaving the tests for completion-at-leisure in the adult studies in Dumfries and Des Moines. The issue is fully discussed in the RPM Manuals (J. Raven et al., 1998a, 1998b, 1998c, 1998d, 1998e) and, more briefly, in an exchange between Gudjonsson (1995) and J. Raven (1995) in *Personality and Individual Differences*.

Before summarizing the available evidence, it is important to note: (1) That, on first contacting the respondents (whose names and addresses had been selected by systematic sampling procedures from almost complete lists of names and addresses), the researchers administered Set I of the APM following the procedures laid down in the Manual. Insofar as it was possible to check the trends documented more fully with the SPM and APM with

TABLE B1
Standard Progressive Matrices: Raw Score Percentiles of 1979 British Standardization among Young People

	Age in years (month)																		
Total score	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)	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)
6		1																	
7	3	1																	
8	4	1																	
9	5	1		1															
10	11	5		3	1	1													
11	13	7	5	4	4	1		1											
12	19	11	9	7	6	2		1				1							
13	25	14	14	9	7	2	3	2				1			1				
14	30	21	20	10	7	3	4	2		1		1	1	1	1				
15	38	28	24	13	8	7	6	2		1		1	1	1	1				
16	49	42	28	16	9	8	7	4		1	1	2	1	1	1				
17	51	50	31	21	9	11	8	5	1	1	2	2	2	1	2				
18	57	58	32	27	11	13	9	6	1	1	2	2	2	1	2				
19	62	63	36	32	13	14	10	7	1	1	2	2	2	1	2				
20	68	66	39	33	16	18	11	8	2	2	3	2	2	1	2				
21	73	69	43	38	17	18	13	8	4	2	3	2	3	2	2	1		1	1
22	75	72	47	40	20	22	15	9	4	3	3	2	3	2	2	1		1	1
23	78	73	54	43	24	27	16	11	4	3	4	3	4	3	2	1		1	1
24	81	78	59	46	26	28	18	12	5	3	6	3	4	3	2	1		1	1
25	82	83	62	50	29	31	18	13	7	4	6	3	5	4	3	2	2	1	1
26	83	84	64	51	34	33	23	15	8	4	6	3	6	4	4	2	2	3	1
27	85	84	68	52	35	36	25	16	9	4	7	4	6	4	4	2	2	3	1
28	86	87	70	55	38	39	27	18	10	6	7	6	6	5	5	2	2	4	2
29	89	89	72	59	42	42	30	20	1	7	7	7	9	5	5	3	2	4	2

30	91	90	77	62	46	43	32	21	14	10	9	7	9	7	5	5	2	5	2
31	91	92	79	64	50	48	34	23	16	15	12	10	9	7	6	5	2	6	4
32	94	93	81	66	54	50	38	26	20	20	15	12	12	9	7	6	3	8	4
33	95	94	83	73	60	56	42	29	24	20	16	13	13	10	8	7	5	9	4
34	96	96	85	77	64	59	46	33	29	23	18	14	15	13	8	8	6	10	5
35	96	97	89	80	70	62	51	35	33	28	23	18	17	16	10	9	8	10	6
36	97	97	93	83	74	67	56	40	37	31	25	21	20	18	10	11	10	14	9
37	97	98	95	85	77	69	60	44	39	35	27	29	23	21	12	12	11	15	11
38	97	99	96	86	81	75	63	50	43	41	31	32	27	23	15	14	13	16	12
39	97	99	97	88	84	78	67	56	51	46	38	34	31	27	19	16	14	19	15
40	98	99	97	91	89	81	73	64	57	51	43	42	38	30	20	21	15	21	19
41	98	100	97	94	92	85	76	69	62	58	51	50	42	37	21	23	20	26	22
42	99		99	95	92	92	80	73	69	62	57	55	51	43	29	29	25	29	24
43	99		100	95	94	93	87	77	73	67	60	64	55	47	35	32	29	35	26
44	100			97	96	95	91	80	80	73	65	67	60	52	43	37	34	39	33
45				97	98	98	93	87	83	77	71	73	70	58	48	45	37	45	39
46				97	98	98	95	91	88	81	78	76	74	62	53	50	41	49	43
47				98	99	100	96	94	90	86	81	79	80	65	60	55	46	53	50
48				98	100		97	95	93	91	84	83	83	69	66	66	51	58	57
49				99			99	97	96	94	90	87	88	73	71	72	58	67	61
50				100			99	98	97	95	93	90	90	78	80	76	63	73	67
51							100	99	98	98	95	93	93	83	84	81	73	76	73
52								100	99	99	97	96	95	87	89	85	79	81	81
53									99	100	98	97	97	91	92	87	83	84	85
54									100		99	97	97	95	94	91	87	87	90
55											99	99	99	98	96	93	93	93	93
56											99	99	100	100	98	96	95	94	94
57											99	100			100	100	96	96	97
58											100						98	97	99
59																	100	100	99
60																			100

this test, the results were similar. (2) That the procedure for administering the RPM laid down in the Manual requires respondents to be allowed to work on their own in a quiet room. Only when the tests are being administered to people who cannot cope with the answer sheets does the tester work through the items individually with the respondent (and this procedure was, in any case, adopted in these adult studies). (3) Respondents were assured of the serious scientific nature of the study (and this was reinforced by the administration of Set I of the APM) and were specifically asked not to seek the collaboration of others.

The most convincing evidence that leaving the tests for completion at leisure can be obtained by looking at Figs. 7 and 8. Most remarkable is the continuity in the shape and level of the curves from the study conducted circa 1942 to 1992. Confidence in the accuracy of both sets of data is enhanced when, as shown in Fig. 8, the results of a study conducted by unrelated researchers halfway between the two Raven studies are superimposed on them. Beyond this, the lack of discrimination among the top 10% of young adults in the 1992 standardization of the test completed at leisure—which might be adduced as evidence that some of the respondents had obtained assistance—had been noted by other researchers (such as Flynn, 1987 and de Lemos, 1989, 1990) who employed the standard administration procedure. Furthermore, the ceiling effect is most noticeable among young adults. It was no greater among those born in 1922 when they were tested in 1992 than it was in 1942, when they were tested using the standard administration procedure. If the clustering of the scores at the top end of the distribution among young adults were due to the extra time available, or to having been able to obtain assistance, it is hard to see why it is not also apparent among older people. What is more, the increase is not apparent in the MHV data (Fig. 9), where respondents could—even more easily—have sought the assistance of a dictionary.

These SPM results are not, however, the only data which bear on the question of what effect leaving test booklets for completion at leisure might have had because data for the APM were collected at the same time. It was therefore possible to compare these data with similar studies conducted elsewhere—such as in the United States and Poland. Unfortunately, these data were either collected from samples (such as Naval recruits) which were known to lack representativeness or from studies [conducted in China (Liu, 1992) or Poland (Jaworowska & Szustrowa, 1991)] in which a time limit had been employed. Nevertheless, the general conclusion was that the effect of leaving the test booklets for completion at leisure cannot have been serious. Since these comparisons were published additional adult data have been collected in Germany (Bulheller & Hacker, 1998). The resulting norms are, in general, well above the British data. Although these high scores are in part attributable to the composition of the sample, they again support the conclusion that the data obtained in the United Kingdom and United States

by leaving the test booklets for completion at leisure cannot be seriously misleading.

There is one final piece of evidence which reinforces the impression that those involved in the Dumfries and Des Moines studies abided by our request not to seek assistance. As Tables MHV7 and MHV12 in J. Raven et al. (1998e) show, the raw scores for the Vocabulary (MHV) test (also completed at leisure in both studies) vary from only 2 above the minimum score, which would have been possible to have obtained, to 2 below the maximum. Had a significant number of people resorted to asking others or using a dictionary, it seems unlikely that such a spread of scores would have been obtained.

REFERENCES

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

Andrich, D., & Dawes, I. (1989). Conversion tables CPM/SPM/APM. In J. Raven & J. H. Court (Eds.), *Research supplement no. 4 to the manual for Raven's Progressive Matrices and Vocabulary Scales* (pp. 23–24). Oxford, England: Oxford Psychologists Press/San Antonio, TX: The Psychological Corporation.

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

Atkinson, J. (1968). *A handbook for interviewers (Government Social Survey).* London: HMSO.

Berelson, B., & Steiner, G. (1964). *Human behavior: An inventory of scientific findings.* New York: Harcourt, Brace & World.

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

Bulheller, S., & Hacker, H. (1998). *Deutsche Bearbeitung und Normierung.* In J. Raven, J. C. Raven, & J. H. Court, Herausgegeben von H. Hacker und S. Belheller, *Manual: Raven's Progressive Matrices und Vocabulary Scales. Teil 4: Advanced Progressive Matrices.* Frankfurt, Germany: Swets Test Services.

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.

Carver, R. P. (1989). Measuring intellectual growth and decline. *Psychological Assessment,* **1**(3), 175–180.

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, November). The use of Raven's Progressive Matrices in Hong Kong: A critical review. *Psychological Test Bulletin,* **2**(2), 40–45.

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. (1998). *A researcher's bibliography for Raven's Progressive Matrices*

and Mill Hill Vocabulary Scales. Obtainable in hard copy and disk format from Mrs. C. J. Raven, 22 Woodplumpton Lane, Broughton, Preston, Lancs. PR3 5JJ, UK.

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.* Oxford, England: Oxford Psychologists Press/San Antonio, TX: The Psychological Corporation.

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, November). The Australian re-standardisation of the Standard Progressive Matrices. *Psychological Test Bulletin,* **2**(2), 17–24.

de Lemos, M. M. (1990, August). The Australian re-standardisation of the Standard Progressive Matrices: A short commentary about the lack of discrimination among the top 10% of 16–18 year olds. *ACER Psychnews,* 9–12.

Deary, I. J. (1993). Inspection time and WAIS-R IQ subtypes: A confirmatory factor analysis study. *Intelligence,* **17,** 223–236.

Deary, I. J. (1995). Auditory inspection time and intelligence: What is the direction of causation? *Development Psychology,* **31,** 237–250.

Deary, I. J. & Stough, C. (1996). Intelligence and inspection time. *American Psychologist, 51*(6), 599–608.

Deming, W. E. (1980). Improvement of quality and productivity through action by management. *National Productivity Review,* **1,** Winter, 12–22.

Duncan, O. D. (1961). A socioeconomic index for all occupations and properties and characteristics of the socioeconomic index. In A. J. Reiss et al. (Eds.), *Occupations and social status.* New York: Free Press.

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 Univ. 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. (1999). Searching for justice: The discovery of IQ gains over time. *American Psychologist,* **54**(1), 5–20.

Foulds, G. A., & Raven, J. C. (1948). Intellectual ability and occupational grade. *Occupational Psychology,* **22,** 197–203.

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.

Gudjonsson, G. H. (1995). Methodological problems with the 1992 standardisation of the SPM. *Personality and Individual Differences,* **18,** 439–442.

Hall, J., & Jones, D. C. (1950). Social grading of occupations. *British Journal of Sociology,* **I,** 31–55.

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

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

Hodge, R. W., Siegel, P. M., & Rossi, P. H. (1964). Occupational prestige in the United States. *American Journal of Sociology,* **70,** 286–302.

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. (Eds.), 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). Oxford, England: Oxford Psychologists Press/San Antonio, TX: The Psychological Corporation.

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.

Horn, J. L. (1994). Theory of fluid and crystallized intelligence. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 443–451). New York: MacMillan.

Hyman, H. H. (1955). *Survey design and analysis: Principles, cases, and procedures.* Glencoe, IL: Free Press.

Inkeles, A., & Rossi, P. H. (1956). National comparisons of occupational prestige. *American Journal of Sociology,* **62,** 329–339.

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.

Kish, L. (1949). A procedure for objective respondent selection within the household. *Journal of the American Statistical Association,* **44,** 380–387. (Reprinted in L. Kish (1965), *Survey sampling*, New York: Wiley.)

Knight, I., & Eldridge, J. (1984). *The heights and weights of adults in Great Britain.* London: HMSO.

Kohn, M. L., Slomczynski, K .M., & Schoenbach, C. (1986). Social stratification and the transmission of values in the family: A cross-national assessment. *Sociological Forum,* **1.**

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

Kratzmeier, H., & Horn, R. (1980). *Manual: Raven-Matrizen-Test. Advanced Progressive Matrices.* Weinheim, Germany: Beltz Test.

Liu, Jin Ming (1992). *Standardisation of the APM in the high schools of Tian Jin.* Report of the Tian Jin Academy of Educational Sciences (in Chinese).

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.

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 Univ. Press.

Matarazzo, J. D. (1990). Psychological assessment versus psychological testing. *American Psychologist, 45,* 999–1017.

Mehlhorn, H. G. (1980). Aspekte der Geistigen Entwicklung Jugendlicher. In W. Freidrich & 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.

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

Newson, J., & Newson, E. (1965). *Infant care in an urban community.* London: Penguin Books.

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

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

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.

Prieler, J. A. (1998, August). *Validation of personnel selection in the Austrian Army.* Paper presented at the International Congress of Applied Psychology Conference, San Francisco.

Raven, J. (1980). *Parents, teachers and children: An evaluation of an educational home visiting programme.* Edinburgh: Scottish Council for Research in Education. Distributed in North America by the Ontario Institute for Studies in Education, Toronto.

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 UK, US, Canada, Germany, and Ireland.* Oxford, England: Oxford Psychologists Press/San Antonio, TX: The Psychological Corporation.

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. (1999). Psychometrics, cognitive ability, and occupational performance. In S. M. Wechsler & R. S. L. Guzzo (Eds.), *Avaliação psicológica: perspectiva internacional* (Psychological assessment: International perspectives) (pp. 299–343). São Paulo, Brazil: Casa de Psicólogo.

Raven, J., et al. (1990/2000). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Research supplement no. 3 (2nd/3rd 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.* Oxford, England: Oxford Psychologists Press/ San Antonio, TX: The Psychological Corporation.

Raven, J., & Court, J. H. (1989). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Research supplement no. 4: Additional national and American norms, and summaries of normative, reliability, and validity studies.* Oxford, England: Oxford Psychologists Press/San Antonio, TX: The Psychological Corporation.

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). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 1: General overview*. Oxford, England: Oxford Psychologists Press/San Antonio, TX: The Psychological Corporation.

Raven, J., Raven, J. C., & Court, J. H. (1998b). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 2: The Coloured Progressive Matrices*. Oxford, England: Oxford Psychologists Press/San Antonio, TX: The Psychological Corporation.

Raven, J., Raven, J. C., & Court, J. H. (1998c). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 3, The Standard Progressive Matrices*. Oxford, England: Oxford Psychologists Press/San Antonio, TX: The Psychological Corporation.

Raven, J., Raven, J. C., & Court, J. H. (1998d). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 4, The Advanced Progressive Matrices*. Oxford, England: Oxford Psychologists Press/San Antonio, TX: The Psychological Corporation.

Raven, J., Raven, J. C., & Court, J. H. (1998e). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 5: The Mill Hill Vocabulary Scale*. Oxford, England: Oxford Psychologists Press/San Antonio, TX: The Psychological Corporation.

Raven, J. C. (1936). *Mental tests used in genetic studies: The performances of related individuals in tests mainly educative and mainly reproductive.* Unpublished master's thesis, University of London.

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

Raven, J. C. (1948). The comparative assessment of intellectual ability. *British Journal of Psychology*, **39,** 12–19.

Raven, J. C. (1965). *Advanced Progressive Matrices: Sets I and II: Plan and use of the scale with a report of experimental work carried out by G. A. Foulds amd A. R. Forbes.* London: H. K. Lewis.

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

Raven, J. C., Styles, I., & Raven, M. A. (1998). *Raven's Progressive Matrices: SPM plus test booklet.* Oxford, England: Oxford Psychologists Press/San Antonio, TX: The Psychological Corporation.

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

Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than *g*. *Journal of Applied Psychology,* **79,** 518–524.

Sahin, N., & Duzen, E. (1994). *Turkish standardisation of Raven's SPM (Ages 6–15).* Paper presented to the 23rd International Conference of Applied Psychology, Madrid.

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.

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

Schaie, K. W., & Strother, C. R. (1968). A cross-sequential study of age changes in cognitive behavior. *Psychological Bulletin,* **70,** 671–680.

Schaie, K. W., & Willis, S. L. (1986). *Adult development and ageing* (2nd edition). Boston: Little, Brown.

Schooler, C., Mulatu, M. S., & Oates, G. (1999). The continuing effects of substantively complex work on the intellectual functioning of older workers. *Psychology and Aging,* **14**(3), 483–506.

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.

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* (Vol. 2, pp. 47–103). Hillsdale, NJ: Erlbaum.

Spearman, C. (1927). *The nature of ''intelligence'' and the principles of cognition* (2nd edition). London, England: Macmillan.

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.* (Report URU-7370). Menlo Park, CA: Stanford Research Institute.

Styles, I., & Andrich, D. (1994, July). *Linking psychometric and cognitive-developmental variables: Quantitative and qualitative aspects.* Invited paper presented at the 23rd International Congress of Applied Psychology, Madrid, Spain.

Styles, I., and Andrich, D. (1997). Faire le lien entre variables psychometriques et variables cognitivo-developpementales regissant le fonctionnement intellectuel. *Psychologie et Psychometrie,* **18**(2/3), 51–69.

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.

Treiman, D. J. (1977). *Occupational prestige in comparative perspective*. New York: Academic Press.

Tuddenham, R. D. (1948). Soldier intelligence in World Wars I and II. *American Psychologist,* **3,** 54–56.

Tuddenham, R. D., Davis, L., Davison, L., & Schindler, R. (1958). *An experimental group version for school children of the Progressive Matrices.* Unpublished manuscript, Univer-

sity 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. A. (1991). Studying intelligence the hard way. *Intelligence,* **15,** 389–395.

Vernon, P. A. (1993). Intelligence and neural efficiency. In D. K. Detterman (Ed.), *Current topics in human intelligence* (Vol. 3, pp. 171–187). Norwood, NJ: Ablex.

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 Univ. 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.

Warner, W. L., & Lunt, P. S. (1941). *The social life of a modern community: Volume I. Yankee city series.* New Haven, CT: Yale Univ. Press.

Warner, W. L., Meeker, M., & Eells, K. (1949). *Social class in America.* Chicago: Science Research Associates.

Watt, D. C. (1998). Lionel Penrose, F.R.S. (1898–1972) and eugenics: Part One. *Notes and Records of the Royal Society of London,* **52,** 137–151.

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, November). Chinese standardisation of Raven's Standard Progressive Matrices. *Psychological Test Bulletin,* **2**(2), 36–39.

(Accepted October 15, 1999)