

The Raven Progressive Matrices: A Review of National Norming Studies and Ethnic and Socioeconomic Variation Within the United States

John Raven
Edinburgh, Scotland

In this paper, some recent results relating to the stability of scores on the Raven Progressive Matrices Test for different subgroups within and between the United Kingdom, the United States, and other Western societies are summarised. Subsequent sections deal with variation over time. A possible explanation for the variation in norms over time and between ethnic groups within the United States is offered.

Raven's Progressive Matrices (RPM) and Mill Hill Vocabulary (MHV) Scales were developed to assess, as straightforwardly as possible, the two components of general intelligence identified by Spearman in 1923 (see Spearman, 1927). These are, respectively, (a) *eductive* ability, that is, the ability to educe correlates, the ability to generate high level schemata, which make it easy to handle complex events, and (b) *reproductive* ability, that is, the ability to recall acquired information.

The available evidence suggests that Raven was successful in developing measures of these abilities. Many researchers have demonstrated that the Progressive Matrices tests are among the purest available measures of *g*, and the Mill Hill Vocabulary Test, which can be administered in a few minutes, correlates well over .9 with full-length intelligence tests (see Raven, Court, & Raven, 1987, for a review of the literature). Nevertheless, within age groups, the Progressive Matrices Test and the Mill Hill Vocabulary Scale correlate only .5. Thus, as Spearman observed, reproductive and eductive abilities would appear to be distinct.

Spearman also noted that, despite their distinct psychological natures, eductive and reproductive abilities interpenetrate and cooperate, contributing cumulatively to effective life performance. Furthermore, the types of activity that people with relatively high eductive ability undertake well are very different from the types of activity that people with high reproductive ability excel in. This finding has not received sufficient attention in educational institutions, which, on the whole, favour students with high reproductive ability (see Hope, 1985; Raven, 1977; Raven, Johnstone, & Varley, 1985).

The Raven Progressive Matrices Tests are available in three forms: the Coloured version (for children), the Standard version (for the entire age range), and the Advanced version (designed to spread the scores of the top 10% on the Standard version). Two versions of the Vocabulary scales are available: the Crichton Vocabulary Scale (for children) and the Mill Hill Vocabulary Scale

(for a cross-section of ability). The MHV scale itself is available in senior and junior forms. Both sets of tests have been revised from time to time over the past 40 years.

Standardisations of the Progressive Matrices

The Progressive Matrices Tests have been used in over 1,600 published psychological studies (see Court, 1988; Court & Raven, 1977 and 1982) and are widely used by applied psychologists.

The Standard Progressive Matrices (SPM) was first fully standardised by Raven on 1,407 children in Ipswich, England, in 1938 (Raven, 1941). Over the years several more normative studies were carried out. The first of these was by Raven himself when he standardised the MHV scale in Colchester, England, in 1943–44 (Raven & Walshaw, 1944). The SPM norms obtained in that study were consistently two raw score points lower than the Ipswich norms. In 1952, Adams reported norms from 11,621 12-year-old children in Surrey, England. The Surrey data were, within the limits of sampling error, very similar to Raven's. Tuddenham et al. (1958), in one of the few studies that attempted to establish the appropriateness or otherwise of the British norms in the United States, tested several school classes of Californian children. They concluded that the British norms were acceptable. In 1963–65, Skanes (see Raven, 1981) tested 4,017 children aged 9½ to 14 in St. Johns, Newfoundland. The similarity between Skanes's results and the 1938 Ipswich norms is striking (Raven, 1981). In 1967, he also tested the entire population of 2,097 Corner Brook, Newfoundland, children aged 10½ to 14½. The results consistently lagged behind the Ipswich norms. In 1972, Byrt and Gill (1973), working with the author, collected data from a nationally representative sample of 3,464 primary school children aged 5½ to 11½ in the Republic of Ireland. The urban norms seemed to correspond to the 1938 Ipswich norms, although the figures for the rural areas lagged behind (Raven, 1981). Up to 1979, therefore, there was apparent uniformity in normative results.

From 1979 on the story began to change. In 1979, Kratzmeier and Horn reported norms from a large German study that were well above those obtained in England in 1938. Melhorn's (1980) East German data were similar. The 1979 British norms appeared to be broadly similar to those obtained in the two German studies (Raven, 1981). Holmes (1980) reported results for British Columbia that were similar, if slightly lower. The New Zealand Council for Educational Research reported closely corresponding results for New Zealand in 1984. Ferjencik (1985) reported data for the Coloured Progressive Matrices for Czechoslovakia that corresponded to a recently reported British study. Work carried out in the United States by Raven, Summers et al. (1986) shows that, although the overall U.S. norms lag behind these international figures, the norms for whites do not. Finally, Zhang (unpublished) has collected data for urban mainland China and shown that, despite the norms collected by Chan (1983) for Hong Kong, they correspond closely with those obtained elsewhere.

The 1979 British Standardisation

We may now examine the 1979 British Standardisation in a little more detail. It was based on a nationally representative sample of over 3,250 children aged 6 to 16.

SPM scores correlated .16 with region of the country, but when the effect of socioeconomic status (SES) was partialled out, the correlation dropped to .07. The correlation between the SPM and SES was .22. However, because age accounted for 46% of the variance, SES accounted for 8.9% of the variance that is not attributable to age. This is equivalent to a within-age correlation between SES and the SPM of .30.

SPM score correlated .68 with age. Thus, more than half the variance was *not* explained by age. One result of this is that the top 10% of 7½-year-olds do better than the bottom 10% of 15½-year-olds. Although these corollaries of the initial observation will be familiar to anyone involved with the measurement of abilities, the first runs counter to the widely held belief that such tests simply measure "intellectual maturity," and the second seems to be largely ignored by the educational system.

As in the 1938 standardisation, item characteristic curve-based (item response theory or Rasch-type) item analyses were carried out separately within each socioeconomic and age group. The items scaled in virtually the same way in all cases. It is therefore difficult to maintain that, in any general sense, the test is foreign to the way of thought of children from certain backgrounds, although that does not mean that it is not unfair to particular children.

U.S. Data

Between 1983 and 1987 some 30 norming studies, involving more than 30,000 students aged 5 to 18 years, were carried out in school districts across the United States of America (Raven, Summers et al, 1986).¹ Each sample was chosen to be representative of the school district from which it was drawn.

The norms that were obtained varied markedly from one school district to another and, within districts, between socioeconomic and ethnic groups. Both ethnicity and SES made independent contributions to the variance. Differences between the norms for school systems catering to white students of differing SES were as great as the ethnic differences within school districts. The ethnic differences seem to correspond to differences in birth weight, infant mortality, and the incidence of serious childhood illness (U.S. Bureau of the Census, 1984).

The data hint that the Hispanic-white difference may be declining: There was no major Hispanic/white difference in the data collected by Stallard (see Raven, Summers et al, 1986) in the Ontario-Montclair school district of California. Yet there had been a marked difference between these two groups in the adjacent county of Riverside when A. R. Jensen (1980) collected his data 15 years earlier. Although Jensen's results have been confirmed in studies conducted in other areas by Burciaga (1973) and Hoffman (1983), Stallard's results are confirmed

in as yet unpublished data recently collected by Felmlee in San Luis valley, Colorado.

Item analyses were run separately for a number of ethnic groups. The results can be indicated briefly by saying that the intercorrelations between the item difficulties established separately for these groups ranged from .98 to 1.00. In addition, Hoffman (1983, 1986) demonstrated that the regression lines of RPM on achievement for the different ethnic groups were parallel but with different intercepts. Thus, although ethnic groups score at different levels on both achievement and matrix tests, the RPM has equal predictive validity within each group.

The Mill Hill Vocabulary Scale

In the 1979 British Standardisation there was, as with the SPM, no variance in MHV scores within region once the effect of SES was partialled out. SES explained 16.2% of the variance not explained by age. MHV scores are, therefore, related more 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, and growth continued at approximately one and a half words per 6-month interval through age 15½. The top 10% of 9-year-olds did better than the bottom 10% of 15½-year-olds.

Separate item analyses were carried out again within eight SES groups. Once more, the reproducibility of the scale properties across groups averaged .97. It would appear to be untrue that children from different backgrounds learn different subsets of dictionary words.

The overall U.S. norms again lagged behind the international data, but the U.S. white norms once more corresponded fairly closely to other cross-cultural data. The test scaled in the same way for (English-speaking) students from different socioeconomic and ethnic backgrounds. This conclusion is similar to that of Deltour (1984) who found that the MHV translated into French with few changes, scaled in the same way, and yielded similar norms.

In sum, this quick and simple measure of reproductive ability has proved to be remarkably robust, and the information it has yielded challenges assumptions commonly made by both psychologists and laymen.

Stability and Change Over Time

The data that have been summarised show that the RPM scales in the same way for children from different socioeconomic backgrounds in both the 1938 and 1979 British standardisations, from different ethnic groups in the 1986 U.S. standardisation, and from different cultures. In addition, the computer-drawn item characteristic curves from the 1979 British standardisation (Raven, 1981) are very similar to the hand-drawn curves from the 1938 standardisation (Raven 1941).

As can be seen from Figure 1, however, the growth curves for British children in 1979 are considerably ahead of those for 1938. (The circled points are those where the unsmoothed 1938 and 1979 data coincide.)

If we can trust the earlier data (and, as we have seen, there is considerable

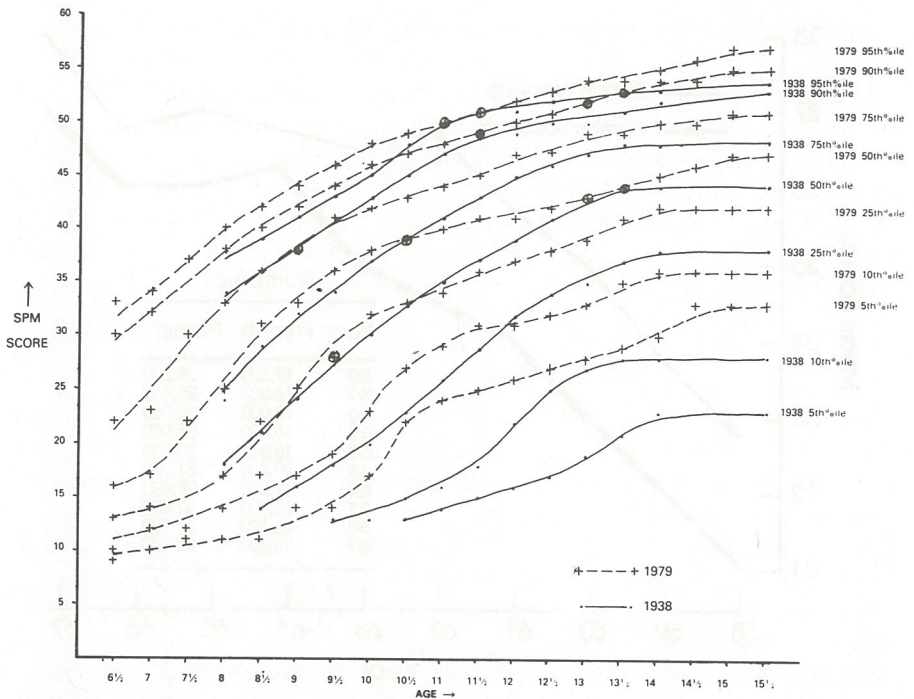


FIGURE 1. Comparison of 1938 and 1979 SPM growth curves

evidence suggesting that we can), it would appear that children now master the abilities tested by the Matrices at an earlier age, and that the scores of the less able plateau above their previous level. However, for the majority of children aged 11½ and over, there has been little increase in score.

Data assembled by Flynn (1987a) both supports and challenges this conclusion. The Progressive Matrices scores of young adults (military conscripts) from a wide variety of westernised societies appear (see, e.g., Bouvier's data in Figure 2) to have been going up steadily—Flynn says “dramatically”—over time. (Flynn cites a round figure of about one standard deviation per generation.²) Flynn suggests that what seems to be a more moderate increase in the United Kingdom—that is, the increase documented in Figure 1—might be explained if the 1938 British norms were too high. This explanation is certainly consistent with the available data. However, an alternative explanation is that British social and welfare provisions were, at that time, well above the international average but have now fallen behind. What is perhaps most striking about Flynn's compendium is that differences between the various norms summarised previously which, taken by themselves, could be dismissed as sampling errors, turn out, in a broader context, to be meaningful.

It would appear from the results summarised earlier that there has been, and still is, considerable—if far from perfect—stability in SPM scores from one Western society to another at any given point in time. However, in common with other test norms (see, e.g., Bouvier, 1969; Garfinkel & Thorndike, 1976;

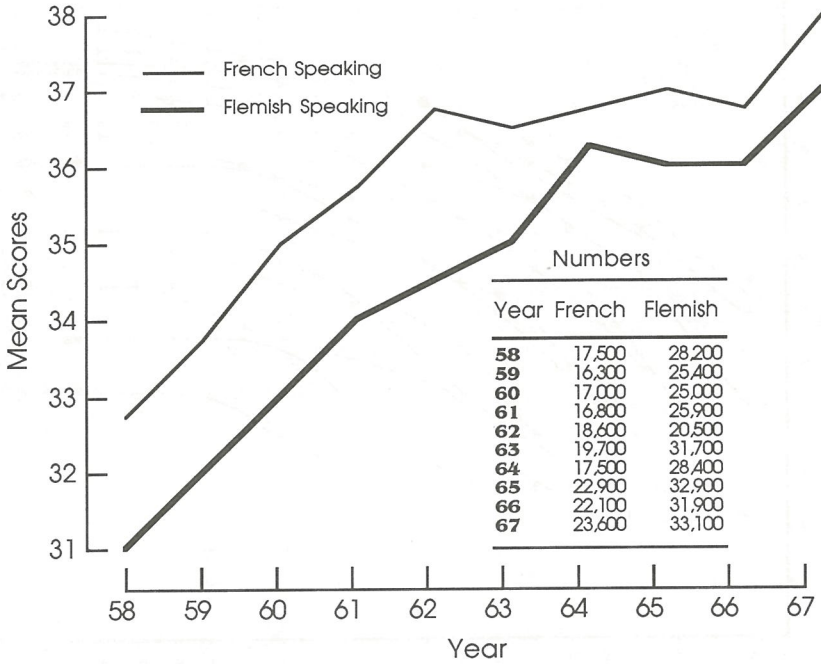


FIGURE 2. Mean SPM scores of Belgian military conscripts (1958–1967) (Redrawn from Bouvier, 1969)

Thorndike, 1975, 1977; and the large number of published and unpublished studies brought together by Flynn, 1984b, 1987b), there has been an increase in mean scores over time, particularly for the less able, at least since 1970 and, according to Flynn, continuously throughout the period (although accelerating more recently).

Thorndike (1977) and Garfinkel and Thorndike (1976) suggest several hypotheses that might explain the increase. However, the data available on the Progressive Matrices do not really support any of them. Thorndike suggests, for example, that the observed acceleration in development may be due to earlier maturity. However, if maturity is a factor, the growth curves for boys and girls should differ more than the evidence summarised in Raven (1981), Raven, Summers et al (1986), and Court (1983) suggests that they do. Thorndike suggests that the differences may be due to the nature of early school education, but the fact that there was little difference between the RPM norms obtained in Scotland and England in the course of the 1979 standardisation suggests that this is unlikely because Scottish infant education remains very formal (HMI, 1980). 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.

Others have suggested that the increases in Progressive Matrices 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 Binet had improved.

Flynn (1987a) (having, in 1984b, queried Thorndike's hypotheses concerning the Binet results) concluded that most of the explanations for the RPM increase that readily come to mind do not hold up. He shows, through a detailed analysis of Leeuw and Meester's (1984) data, that changes in *level* of education could account for only one of the 20 IQ-point gains in RPM scores documented among servicemen. Changes in the intellectual quality of the home environment, as indexed by SES, could account for little more.

The variation in mean scores between ethnic groups within the United States does seem to correspond to variation between the same groups in height, birth weight, and infant mortality. Height and birth weight have, like intelligence test scores, increased over the past 80 years.³ The most probable explanation of the increase in scores over time therefore seems to be that it is due to the same variables as increases in height and birth weight and decline in infant mortality—that is, to improved nutrition, welfare, and hygiene. What it is about these variables that is important is as obscure for height and birth weight as it is for intellectual ability. However, the fact that such variables do have important effects on RPM score as well as birth weight and height is shown in a remarkable study carried out in Aberdeen, Scotland, most of the results from which have never been published (Baird & Scott, 1953; Scott, Illsley, & Thomson, 1956). In this study, calcium intake was used as an index of quality of diet, and it was shown that this has a marked impact on all three of the outcomes mentioned, and that the relationship held both within and between socioeconomic groups. More recently, Benton and Roberts (1988) have shown that vitamin and mineral supplements appear to have a rapid and marked effect on eductive, but not reproductive, ability.

The Origins of Variance

It follows from the results just summarised that, despite their persistence over time (see also the persistence of the French-Flemish speaking difference in Belgium shown in Figure 2), the differences between ethnic groups within the United States cannot be regarded as immutable: Most of the current ethnic norms in the United States lie between the 1938 and 1979 British norms, and, as we have seen, there is some evidence that the Hispanic-white difference may be disappearing. (The significance of the increases over time for the arguments pointing to a genetic explanation of the ethnic differences have been spelled out by Flynn, 1987b).

There have been a number of empirical studies of the factors that increase or decrease RPM scores. 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 that influence the development of eductive ability are not the obvious cultural and socioeconomic variables that 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. The development of the ability to perceive and think clearly appears to be promoted by "democratic" childrearing practices in the home and by "open" educational activities in the school (Chan, 1981; McGillicuddy-DeLisi, DeLisi, Flaughner, & Sigel, 1987; Miller, Kohn, & Schooler, 1985; Raven, 1980; Raven et al., 1985; Sigel, 1986; Sigel & McGillicuddy-DeLisi, 1984; Stallings & Kaskowitz, 1974). It is also promoted by work that demands high levels of problem solving and responsibility (Jaques, 1976; Kohn & Schooler, 1978; Lempert, 1986). However, it appears that the child-rearing and educational practices that increase Progressive Matrices scores *depress* reading, writing, and arithmetical ability if these abilities are measured by conventional educational tests that load heavily on reproductive ability (Sigel, 1986; Stallings & Kaskowitz, 1974).⁴

Experimental attempts to teach the strategies required to solve matrix problems (e.g., Budoff & Corman, 1976; Budoff, Corman, & Gimon, 1976; Feuerstein, 1979, 1980; Haywood et al. 1982; Jacobs, 1977; Wortman, 1968) yield dramatic short-term results, but it is not clear that these affect educative ability in any basic sense (Savell, Twohig, & Rachford, 1986). What is notable about at least some "democratic" parents and "open" educators is that they extend children's utterances and encourage them to ask questions and to invent ways of perceiving and conceptualising things for themselves. They also encourage them to share in their own problematizing and thinking about the long-term consequences of their actions in the context of their own personal views of how society works.⁵ In this context it is, however, important to note that attempts to introduce such activities into schools in general have not succeeded even in getting teachers to change their behaviour (Fraleigh, 1981; HMI 1980; Raven et al., 1985). The same has been true of attempts to get parents to change their behaviour (Raven, 1980). This is almost certainly why overall evaluations of intervention programmes such as Head Start and Follow-Through (e.g., Bock, Stebbins, & Proper, 1977; Spitz, 1986) show no effect.

The evidence that child-rearing practices can influence scores suggests that there is possibly a more psychologically based explanation of the increase in RPM scores over time than the explanation in terms of nutrition and hygiene favoured earlier. This is that the move toward smaller families, combined with welfare provisions that satisfy the more fundamental needs in Maslow's (1954) hierarchy, has permitted more parents to adopt child-rearing practices that facilitate their children's cognitive development. These parents may encourage their children to think things out for themselves; to share in their own thought processes and, in particular, to think about the operation of complex social processes (society) and consider the probable long-term, and as yet intangible, social consequences of their actions. If they do, these experiences might well be reflected in increases in both cognitive ability and the development of internalised codes to guide behaviour. When these more socialised children reach school, their teachers can change their discipline strategies and treat them in a more developmental way. And when they get to work these young people might find that it, too, has become more cognitive, requiring them to take more responsibility for the personal, organisational, and societal consequences of their actions. If

work has changed in this way, Jaques (1976), Kohn and Schooler (1978), and Lempert (1986) already have demonstrated that it further promotes the development of problem-solving ability. What is important about this possible explanation of the observed increase in scores is that it is not disconfirmed by the observations that earlier led us to reject Thorndike's hypotheses.

It is important to conclude this section on a less "environmentalist" note by juxtaposing the fact that Progressive Matrices scores can be influenced by and have changed over time with the fact that the bulk of the variance is still between children from similar ethnic and socioeconomic backgrounds. Indeed, as Maxwell (1969) and Jencks et al. (1973) have shown, two-thirds of the variance in intelligence test scores is between children from the same families.

Concluding Comments

The first general conclusion that may be drawn out of this material is that it does not really support Flynn's (1987a) contention that "IQ tests do not measure intelligence but rather a correlate with a weak causal link to intelligence." It is easiest to demonstrate this by citing parallel data on height. According to Flynn (personal communication), height has, since well before the turn of the century, been increasing at about two thirds of a standard deviation per generation. This is of the same order of magnitude as the one standard deviation per generation figure he cites for RPM scores. For height, too, sex and ethnic differences abound, and the differences persist despite intergenerational increases. Neither the intergenerational increases nor the sex and ethnic differences imply that rulers cannot validly be used to measure height. Flynn buttresses his conclusion that IQ scores have little meaning by suggesting that, if people's intelligence really has improved, more patents should have been filed. This has not happened. But no one would suggest that a prerequisite to rulers being regarded as valid measures of height would be that the number of Olympic high jump gold medalists should have gone up. (Incidentally, would not the number of books and articles published be a better criterion against which to judge the construct validity of IQ tests than the number of patents filed? The number has certainly increased over time.)

My own view is that it would have been virtually impossible to have obtained meaningful evidence on the issues in question if the tests had not been available. The impressive stability in the norms across cultures at any given point in time, the stability in the test properties and item statistics across time and socioeconomic and ethnic group, the consistency of the increases from year to year, and the persistence of the ethnic differences imply that we are dealing with psychological processes of fundamental importance.

Despite my disagreement with Flynn's general conclusion, I feel that his compilation of data is of the greatest theoretical and practical importance. These data demonstrate that some, as yet unidentified, features of the environment have dramatic effects on educative ability.

Another set of insights that stem from the material summarised in this article are those that relate to the controversy about test bias which has troubled American education for two decades. Because RPM and MHV scores are

relatively unambiguous, the data that have been reviewed direct attention toward a constructive search for explanations of differences and for educational policies that reflect the socioeconomic realities, values, and educational needs of the groups concerned. (See Raven (1987) and the article by Tharp, Gallimore, and others (Tharp et al., 1984) for a fuller discussion the way in which this might be done.) Nevertheless, to fully capitalize on these observations it is essential for psychologists to undertake the fundamental research that is needed to develop measures of a wider range of human talents. It seems that we are still trying to administer our educational system with the aid of concepts and tests developed by Binet at the turn of the century. (A new attempt to provide a conceptual framework and measurement model appropriate to thinking about the assessment, development and release of a wider range of talents is available in Raven, 1984, 1988, in press-a.)

It may be useful to expand on what was said in the last paragraph by taking an example. The material that has been presented shows that, if Federal US funds for Gifted Education were administered on the basis of National US norms, those funds would be channelled into what are already very wealthy school districts. If this is regarded as inequitable, and if it is therefore felt that local instead of national norms should be used to determine the level of provision of, and eligibility for entry to, both "gifted" and "remedial" forms of special education, then the logic of the argument points to the use of local ethnic norms. In the end, therefore, decisions about which norms to use depends on judgments about whether Special Education is a good thing or a bad thing: does it help or does it label? By introducing these facts and considerations into the debate it is possible to direct attention to the goals of policy and the steps that are required to reach them.

Psychologists should be arguing much more forcefully for a major role in identifying policy goals and the processes to be used to reach them and in developing the tools needed to administer the programmes and assess their effects. If they did this, they would find themselves developing a whole new range of understandings of the qualities which are to be fostered in the course of educational programmes, the processes to be used to nurture them, and the psychometric models to be used to assess educational processes and outcomes (Raven, 1977, 1984, 1988).

Summary

At its most basic level, the data briefly summarised in this paper suggest that there is considerable stability in SPM and MHV performance both within and between societies with a literary tradition at any given point in time. This is true at the level of item statistics as well as both mean scores and variance. On the other hand, there has been an impressively continuous increase in scores over time. Despite these increases, the differences between socioeconomic and ethnic groups, both within the United States and across cultures, remain.

From a more fundamental point of view the material suggests that the use of theoretically based tests which have good psychometric properties has:

- demonstrated that we are dealing with psychological processes that are of

fundamental importance. The reproducibility of the scale across time and socioeconomic and ethnic groups, the stability of the norms across cultures, the persistence of ethnic differences, and the consistency of the rate of increase all support this conclusion.

- shown that it is not true that all human abilities have basically the same patterns of intercorrelation with independent and dependent variables. This is important because it opens the way to solutions to social and educational problems which are not available to those who have uncritically accepted the belief that the bulk of the variance in human performance is attributable to a single factor.

- made it possible to redirect efforts to find socially acceptable solutions to the socioeducational problems that have been associated with previous attempts to use psychological concepts and tools for the enormous range of abilities and talents that are present in any school class.

- contributed significantly, if in unexpected ways, to answering some of the highly contentious and politicised research questions that led to these tests being developed in the first place. The use of the same test in a wide variety of studies, in field surveys, laboratory experiments, and field trials, in many cultures, and over an extended period of time has, on the one hand, demonstrated that the environment can have a previously unsuspected impact on educative ability and, on the other, shown that most of the hypotheses commonly advanced to explain the increase in scores over time are invalid. Firm evidence of an increase over time has eluded those who have employed tests of General Intelligence that sample a wide range of content and process because these tests include items that measure knowledge which is either culturally specific or which becomes increasingly well-known or obsolete. The massive effect of the environment has also eluded researchers who have worked with twins. But the results are not only disconcerting to those who believe that *g* is largely inherited. They also pose problems for those who are inclined to believe that socioeconomic and educational environments markedly influence intellectual performance, and particularly those who seek to explain low scores by reference to a cultural-deficit hypothesis.

The available evidence suggests that the factors responsible for both the increase over time and the ethnic differences will be found to be among those that account both for increases over time and ethnic differences in height, birth weight, and incidence of infant mortality. However, the evidence is far from conclusive and more psychologically oriented hypotheses are still tenable.

Notes

¹Compilation of these norms is continuing. The author would welcome correspondence from anyone able to contribute to the process and, in particular, from anyone able to establish norms for an urban, black sample.

²There is no doubt that there has been a major increase in scores over time and that Flynn has done an outstanding service by unearthing and bringing together studies that demonstrate this. Likewise, there is no doubt it was necessary for Flynn, like most other researchers, to employ statistics that do not fully satisfy the relevant statistical assump-

tions in order to make progress. Nevertheless, it is important to recognise that expression of the amount of the increase in IQ units gives an unjustified impression of precision. The problems inherent in the use of IQs are fully discussed in Raven, Summers et al. (1986) and cannot be discussed here. The easiest way to hint at them is by taking an example. The within-age distributions of raw scores on many tests deviate markedly from a Gaussian ("normal") curve. They are frequently both bimodal and have tails that do not correspond to the shape of the Gaussian curve. The result is that the IQ of the same child, on the same test, judged against the same standardisation sample, can be 40 if the statistician who calculated the norms made one set of assumptions and 65 if he or she made other assumptions. It is important to recognise that these technicalities mean that it would be inappropriate for researchers to embark on a debate about the precise magnitude of the rate of increase.

³Although the Progressive Matrices data are limited to 50 years, Tuddenham's (1948) army data go back to 1914 and the Binet data earlier still (Thorndike, 1975).

⁴It is not known what the results would be if these academic abilities were measured by tests that take into account the use of such abilities to find information one wants, communicate what is important to one, and lead one's life effectively. Raven et al. (1985), and especially Raven (1988b, in press), discuss the practical implications of these limitations of conventional educational tests.

⁵For fuller discussions see Raven (1987, in press-b, in press-c, in press-d), but the main references are Tough (1973), McGillicuddy-DeLisi (1985), Sigel and Kelley (1988), Gallimore, Boggs, and Jordan (1974), Tharp and Gallimore (1984), Raven (1980), and Raven et al. (1985).

References

- Adams, A. E. (1952). *Analysis of Raven's matrices scores: Preliminary report*. Surrey, England: Surrey Educational Research Association.
- Baird, D., & Scott, E. M. (1953). Intelligence and childbearing. *Eugenics Review*, 45, 139-154.
- Benton, D., & Roberts, G. (1988). Effect of vitamin and mineral supplementation on intelligence of a sample of schoolchildren. *Lancet*, 140-143.
- Bock, G., Stebbins, L. B., & Proper, E. C. (1977). *Education as experimentation: A planned variation model (Vol. IV-B). Effects of follow through models*. Cambridge, MA: Abt Associates.
- Bouvier U. (1969). *Evolution des cotes a quelques test*. Belgium: Centre de Recherches, Forces Armees Belges.
- Budoff, M., & Corman, L. (1976). Effectiveness of a learning potential procedure in improving problem-solving skills of retarded and non-retarded children. *American Journal of Mental Deficiency*, 81, 260-264.
- Budoff, M., Corman, L., & Gimon, A. (1976). An educational test of learning potential assessment with Spanish speaking youth. *Interamerican Journal of Psychology*, 10, 13-24.
- Burciaga, L. E. (1973). *A research study on the Raven Coloured Progressive Matrices among school children of the El Paso public schools*. Doctoral dissertation, University of Texas, El Paso.
- Byrt, E., & Gill, P. E. (1973). Standardisation of Raven's Standard Progressive Matrices and Mill Hill Vocabulary for the Irish population ages 6-12. Master's thesis, National University of Ireland, University College Cork.
- Chan, J. (1981). Correlates of parent-child interaction and certain psychological variables

- among adolescents in Hong Kong. In J. L. M. Binnie-Dawson (Ed.), *Perspectives in Asian cross-cultural psychology*. Lisse, Netherlands: Zwets and Zeitlinger.
- Chan, J. (1983). Norms for Hong Kong. In J. C. Raven, J. H. Court, & J. Raven. *A manual for Raven's progressive matrices and vocabulary tests* (1983 ed. "Standard Progressive Matrices" Section, Table SPM XII). London: Lewis.
- Court, J. H. (1983). Sex differences on Raven's Progressive Matrices: A review. *Alberta Journal of Educational Research*, 29, 54-74.
- Court, J. H. (1988). *A researcher's bibliography for Raven's Progressive Matrices and Mill Hill Vocabulary scales* (7th ed.). Published by J. H. Court, 500 Goodwood Rd., Cumberland Park, South Australia 5041.
- Court, J. H., & Raven, J. (1977 & 1982). Summaries of reliability, validity, and normative studies for Raven's Progressive Matrices and Vocabulary Scales. In *A manual for Raven's progressive matrices and Mill Hill vocabulary scales* (Research Supp. No. 2). London: H. K. Lewis; San Antonio, Texas: The Psychological Corporation.
- Deltour, J. J. (1984). *Une nouvelle methode d'evaluation rapide de la deterioration mentale*. Unpublished manuscript, University of Liege, Laboratoire de Psychologie Appliquee.
- Ferjencik, J. (1985). *Manual: Coloured Progressive Matrices*. Bratislava, Czechoslovakia: Psychodiagnostické, a Didaktické Testy.
- Feuerstein, R. (1979). *The dynamic assessment of retarded performers*. Baltimore, MD: University Park Press.
- Feuerstein, R. (1980). *Instrumental enrichment*. Baltimore, MD: University Park Press.
- Flynn, J. R. (1984a). IQ gains and the Binet decrements. *Journal of Educational Measurement*, 21, 283-290.
- Flynn, J. R. (1984b). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, 95, 29-51.
- Flynn, J. R. (1987a). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, 101(2).
- Flynn, J. R. (1987b). Race and IQ: Jensen's case refuted. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy* (pp. 221-232). Lewes, England: Falmer Press.
- Fraley, A. (1981). *Schooling and innovation: The rhetoric and the reality*. New York: Tyler Gibson.
- Gallimore, R., Boggs, J. W., & Jordan, C. (1974). *Culture, behavior and education*. Beverly Hills, CA: Sage.
- Garfinkel, R., & Thorndike, R. L. (1976). Binet item difficulty: Then and now. *Child Development*, 47, 959-965.
- Haywood, H. C., & Arbitman-Smith, R. (1981). Modification of cognitive functions in slow-learning adolescents. In P. Mittler (Ed.), *Frontiers of knowledge in mental retardation: Vol 1. Social, educational, and behavioural aspects*, (pp. 129-140). Baltimore, MD: University Park Press.
- Haywood, H. C., Arbitman-Smith, R., Bransford, J. D., Towery, V. R., Hannel, I. L., & Hannel, M. V. (1982). *Cognitive education with adolescents: Evaluation of instrumental enrichment*. Paper presented at the International Association for the Study of Mental Deficiency.
- Her Majesty's Inspectors (Scotland) (1980). *Learning and teaching in primary 4 and primary 7*. Edinburgh: Her Majesty's Stationery Office.
- Hoffman, H. V. (1983). *Regression analysis of test bias in the Raven's Progressive*

- Matrices for Anglos and Mexican-Americans*. Unpublished doctoral dissertation, University of Arizona, Department of Educational Psychology.
- Hoffman, H. V. (1986). In J. Raven, J. H. Court, & J. Raven, *Manual for Raven's progressive matrices and vocabulary tests. (A compendium of North American normative and validity studies*, Research Supp. No. 3). London: H. K. Lewis.
- 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*. University of British Columbia: Faculty of Education.
- Hope, K. (1985). *As others see us: Schooling and social mobility in Scotland and the United States*. New York: Cambridge University Press.
- Jacobs, P. I. (1977). *Up the IQ!* New York: Wyden Books.
- Jaques, E. (1976). *A general theory of bureaucracy*. London: Heinemann.
- Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B., & Michelson, S. (1973). *Inequality: A reassessment of the effect of family and schooling in America*. New York: Basic Books; London: Penguin.
- Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.
- Jensen, M. R. (1985). The LPAD and low functioning children. In *To sing our own songs* (Report of a Workshop for American Indian Educators) New York: Association on American Indian Affairs.
- Kohn, M. L., & Schooler, C. (1978). The reciprocal effects of the substantive complexity of work and intellectual flexibility. A longitudinal assessment. *American Journal of Sociology*, 84, 24-52.
- Kratzmeier, H., & Horn, R. (1979). *Manual: Raven-Matrizen-Test. Standard progressive matrices*. Weinheim, West Germany: Beltz Test.
- Leeuw, J. de, & Meester, A. C. (1984). Over het intelligente-onderzoek 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.
- Lempert, W. (1986). Sozialisation und Persönlichkeitsbildung in beruflichen Schulen, dargestellt am Beispiel der Entwicklung moralischer Orientierung. *Die berufsbildende Schule*, 38, 148-160.
- Maslow, A. H. (1954). *Motivation and personality*. New York: Harper.
- Maxwell, J. N. (1969). *Sixteen years on*. Edinburgh: Scottish Council for Research in Education.
- 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*. Hillsdale, NJ: Erlbaum.
- McGillicuddy-DeLisi, A. V., DeLisi, R., Flaugh, 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*. New York: Cambridge University Press.
- Melhorn, H. G. (1980). Aspekte der Geistigen Entwicklung Jugendlicher. In W. Freidrich & H. Muller (Eds.), *Zur Psychologie der 12 bis 22 Jahrigen*. East Berlin: VEB Deutscher Verlag der Wissenschaften.
- Miller, K. A., Kohn, M. L., & Schooler, C. (1985). Educational self-direction and the cognitive functioning of students. *Social Forces*, 63, 923-944.
- New Zealand Council for Educational Research. (1984). *Standard Progressive Matrices: New Zealand norms supplement*. Wellington: Author.
- Raven, J. (1977). *Education, values and society: The objectives of education and the nature and development of competence*. London: H. K. Lewis.

- Raven, J. (1980). *Parents, teachers and children*. Edinburgh: Scottish Council for Research in Education.
- Raven, J. (1981). 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. In J. C. Raven, J. H. Court, & J. Raven (1987), *A manual for Raven's progressive matrices and vocabulary tests* (Research Supp. No. 1). London: H. K. Lewis.
- Raven, J. (1984). *Competence in modern society: Its identification, development and release*. London: H. K. Lewis.
- Raven, J. (1987). Values, diversity and cognitive development. *Teachers College Record*, 89, 21–38.
- Raven, J. (1988). The assessment of competencies. In H. D. Black & W. B. Dockrell (Eds.), *New developments in educational assessment: British Journal of Educational Psychology, Monograph Series No. 3*, 98–126.
- Raven, J. (in press-a). A model of competence, motivation and its assessment. In H. Berlak (Ed.), *Assessing academic achievement: Issues and problems*. Madison, WI: National Center for Effective Secondary Schools.
- Raven, J. (in press-b). Equity in diversity: The problems posed by values—and their resolution. In F. Macleod (Ed.), *Families and schools: Issues in accountability and parent power*. Brighton, England: Falmer Press.
- Raven, J. (in press-c). Parents and schools in early childhood. *British Journal of Educational Psychology* (Special issue).
- Raven, J. (in press-d). *Progressive matrices: Solution strategies and their implications for education*. Madrid: Ministry of Education.
- Raven J., & Court, J. H. (in press). *Additional normative, reliability, and validity data—to 1988*. Research supplement No. 4 to Raven, J. C., Court, J. H., & Raven, J. *Manuel for Raven's Progressive Matrices and Vocabulary Scales*. London: H. K. Lewis. San Antonio, TX: Psychological Corporation.
- Raven, J., Johnstone, J., & Varley, T. (1985). *Opening the primary classroom*. Edinburgh: Scottish Council for Research in Education.
- Raven, J., Summers, W. A., et al. (1986). A compendium of North American normative and validity studies. In J. C. Raven, J. H. Court, & J. Raven, *A manual for Raven's progressive matrices and vocabulary tests* (Research Supp. No. 3). London: H. K. Lewis; San Antonio, Texas: The Psychological Corporation.
- Raven, J. C. (1941). Standardisation of progressive matrices. *British Journal of Medical Psychology*, 19, 137–150.
- Raven, J. C., Court, J. H., & Raven, J. (1987). *A manual for Raven's Progressive Matrices and Vocabulary Tests*. London: H. K. Lewis; San Antonio, Texas: The Psychological Corporation.
- Raven, J. C., & Walshaw, J. B. (1944). Vocabulary Tests. *British Journal of Medical Psychology*, 20, 185–194.
- Savell, J. M., Twohig, P. T., & Rachford, D. L. (1986). Empirical status of Feuerstein's "Instrumental Enrichment" (FIE) Technique as a method of teaching thinking skills. *Review of Educational Research*, 56, 381–409.
- Scott, E. M., Illsley, R., & Thomson, A. M. (1956). A psychological investigation of primigravidae: II. Maternal social class, age, physique and intelligence. *Journal of Obstetrics and Gynaecology of the British Empire*, 63, 338–343.
- Sigel, I. E. (1986). Early social experience and the development of representational competence. In W. Fowler (Ed.), *Early experience and the development of competence. New Directions for child development* (No. 32). San Francisco: Jossey-Bass.

- Sigel, I. E., & Kelley, T. D. (1988). A cognitive developmental approach to questioning. In J. Dillon (Ed.), *Classroom questioning and discussion: A multidisciplinary study*. Norwood, NJ: Ablex.
- Sigel, I. E., & McGillicuddy-DeLisi, A. V. (1984). Parents as teachers of their children: A distancing behaviour model. In A. D. Pellegrini & T. D. Yawkey (Eds.), *The development of oral and written language in social contexts*. Norwood, NJ: Ablex.
- Skanes, G. R., Sullivan, A. M., Rowe, E. J., & Shannon, E. (1974). Intelligence and transfer. *Journal Educational Psychology*, 66, 563–568.
- Spearman, C. (1927). *The nature of "intelligence" and the principles of cognition* (2nd ed.). London: Macmillan.
- Spitz, H. H. (1986). *The raising of intelligence: A selected history of attempts to raise retarded intelligence*. Hillsdale, NJ: Erlbaum.
- Stallings, J., & Kaskowitz, D. (1974). *Follow through classroom observation evaluation 1972–1973*. (Report No. URU-7370). Menlo Park, CA: Stanford Research Institute.
- Tharp, R. G., Jordan, C., Speidel, G. E., Au, K. H. P., Klein, T. W., Calkins, R. P., Sloat, K. C. M., & Gallimore, R. (1984). Product and process in applied developmental research: Education and the children of a minority. In M. E. Lamb, A. L. Brown, & B. Rogoff (Eds.), *Advances in developmental psychology* (Vol. 3, pp. 91–144). Hillsdale, NJ: Lawrence Erlbaum.
- Thorndike, R. L. (1975). *Mr. Binet's test 70 years later*. Presidential address presented at the annual meeting of the American Educational Research Association,
- Thorndike, R. L. (1977). Causation of Binet IQ decrements. *Journal of Educational Measurement*, 14, 197–202.
- Tough, J. (1973). *Focus on meaning: Talking with some purpose to young children*. London: Allen and Unwin.
- Tuddenham, R. D. (1948). Soldier intelligence in World Wars I and II. *American Psychologist*, 3, 54–56.
- Tuddenham, R. D., Davils, L., Davison, L., & Schnindler, R. (1958). *An experimental group version for school children of the progressive matrices*. University of California (Mimeo). (See Abstract: *J Consult. Psychol.*, 22, 30.)
- United States Bureau of the Census. (1984). *Statistical abstract of the United States, 1983*. Washington, DC: U.S. Government Printing Office.
- Wortman, R. A. (1968). *Coaching and teaching in retardates: The Raven Matrices as a learning situation* (Final Report, Project 6–8441). Washington, DC: U.S. Department of Health, Education, and Welfare,

Author

JOHN RAVEN, Consultant, 30 Great King St., Edinburgh EH3 6QH, Scotland. Degrees: BSc, University of Aberdeen; Dip. Soc. Psych., London School of Economics; PhD, University of Dublin. Specializations: broadly based educational evaluation, links between education and society.