# NATURE VERSUS NURTURE : METHODOLOGICAL QUESTION AND FISHERIAN ANSWER

Rian van Blokland - Vogelesang

Department of Data Theory
University of Leiden

RR-92-05

# NATURE VERSUS NURTURE : METHODOLOGICAL QUESTION AND FISHERIAN ANSWER

**Rian van Blokland - Vogelesang**

**Department of Data Theory**

**University of Leiden**

# NATURE VERSUS NURTURE: METHODOLOGICAL QUESTION AND FISHERIAN ANSWER

Rian van Blokland-Vogelesang
University of Leiden, Department of Data Theory
P.O. Box 9555, 2300 RB Leiden, The Netherlands

The Nature-Nurture controversy is mostly discussed by using results from correlational studies or analyses of covariance. By using these techniques, two distinct effects are disregarded. First, a difference in means is overlooked: the additive effect is lost by computing the Pearson correlation between the variables under study (Father's intelligence and Son's intelligence, in this case). Secondly, a longitudinal effect is overlooked: a continuing increase in the mean level of intelligence in the population since the thirties. Many phenomena expressed in terms of correlation can be more clearly thought of in terms of additive factors, just as Fisher did. In accordance with this approach, we investigate a set of data on Father-Son pairs on the Raven Progressive Matrices Test using analysis of variance and variance components analysis. Results are compared with previous results from path analysis for this data. Comparable studies for America and Japan are discussed also. It is argued that the increase in IQ is due to environmental changes since the thirties: an effect of Nurture. Besides the large Nurture effect, evidence is given for the existence of a Nature effect.

## 1 Introduction

The Nature-Nurture controversy is mostly discussed by using results from correlational studies and analyses of covariance. If we calculate Pearson product moment correlations between IQ values in two matched groups (say, Fathers and their Sons), then for each variable (say, Father-IQ and Son-IQ), the average value is subtracted from each observation. This implies a *change of scale* for both variables: further analyses cannot anymore concern differences in average values between the two variables. Therefore, an additive effect, the average IQ-difference between the two groups, is lost by calculating correlations between the IQ-variables under study. The same holds true if covariances are calculated. In contrast, in performing an analysis of variance, the total variance in IQ is partitioned into independent contributions by the two groups. This *additive* effect is thus calculated from the average group-difference, the so-called between groups variation, and is tested against the residual variance or error variance. The residual variance is calculated from the within-groups variation. Therefore, variance-based analyses are explicitly designed to evaluate differences between group-means.

To Charles Spearman (1863 - 1945) we owe the concept of a general factor, $g$, for intelligence. The $g$-factor is the essence of what most people think of as "intelligence" (Jensen, 1973, p. 77; 1981, p. 59). Highly $g$-loaded tests show a greater number of relatively high correlations with other tests, while the least $g$-loaded tests show only small correlations with other tests. The $g$-factor is most clearly manifested in items calling for inductive and deductive reasoning and abstraction, and in items that require mental manipulation of images, symbols, words, numbers, or concepts. The Raven Progressive Matrices Test (Jensen, 1981, p. 97, 210, 301-302) is an example of a test with a high loading on Spearman's general ability factor, $g$. The Raven Test calls for perceiving key features and relationships among simple geometric figures and designs, and discovering the rules that govern the differences among the elements in the matrix. Furthermore, this test is 'culture-fair' to a high degree (Jensen, 1973, p. 97).

In accordance with Fisher's (1925, 1990) approach, we present results on Father-Son pairs with respect to the Raven Progressive Matrices Test, using analysis of variance and variance components analysis (the multi-level model, see Goldstein, 1987; Aitkin and Longford, 1986) in order to investigate the possible contributions of *Nature* (hereditary factors) and *Nurture* (environmental variables), respectively. Our conclusion is compared with two other analyses of this data. The first description is Vroon, De Leeuw, and Meester (1986) who use covariance structure analysis to disentangle hereditary and environmental factors. They were not able to find evidence either for a nature effect, nor for a nurture effect. The second description is Van den Berg (1989a, b), who just shows the large increase in Sons' Raven scores compared to Fathers' Raven scores, and describes environmental factors that might have contributed to this enormous increase.

In section 2, two controversies Fisher and Pearson were involved with (see Fisher Box, 1978; Gower, 1991) are treated. These controversies concern first the size of correlations to be expected for hereditary factors and, secondly, the evaluation of genetic differences by way of analysis of variance as distinguished from correlational analysis. In section 3, these methodological differences are illustrated with reference to the Skodak-Skeels (1949) study on adopted children. Section 4 shows the analysis of differences between Dutch Father-Son pairs on the Raven Progressive Matrices Test; the difference in Raven score between Fathers and their Sons is considerable (Van den Berg, 1989a, b), so we expect an additive effect resulting from quickly changing environmental and cultural conditions, and, in particular, from improved and prolonged *education*. Comparable results for America and Japan are reported by Flynn (1984) and Anderson (1982). The intensity of the additive effect might be changing with the years, which is to say that there might be a *longitudinal* effect. To measure this longitudinal effect, we condition on Year of Father's birth, since quality and amount of schooling is tied to year of birth for most of the people. In section 5, further evidence for the Raven effect is given, and in section 6, a previous study of the Father-Son data (Vroon et al., 1986) is criticized. Section 7 closes with a discussion.

## 2 The Fisher-Pearson Controversies

There are two separate controversies with which Fisher was involved. These are: (a) his major work reconciling the genetics of continuous traits with the discrete nature of Mendelian genetics (see Fisher, 1922), and (b) the controversy of replacing Pearson's inter- and intra-class correlation methodology by additive models with their analysis-of-variance summary. Both were meant to handle the possibility of different mean values between classes, but Pearson's approach becomes hopelessly complicated, except in the most simple cases.

The work of abbé Gregor Mendel had established the fact of discrete biparental inheritance of a number of factors (genes) in the garden pea: the plants were tall or short, the seeds smooth or wrinkled; and these genes acted independently of each other. Mendel's theory is based on combinatorics and explains the discontinuities of inheritance, whereas Darwin's theory implies a continuous evolution, showing a progressive series of adaptive improvements. An example of this is the phenomenon of 'twining' of plants, which enables a plant to reach in one season a height where its leaves are more exposed to sunlight and air, without having to develop a thick woody trunk. Twining is therefore an adaptation, which is improved in some species by the conversion of leaves into tendrils and even further in others where the tips of the tendrils when they strike a solid surface are converted into adhesive disks. Biometrics, as developed by Pearson, was based on Darwin's theory of continuous evolution. Pearson was one of the champions of Darwin's theory.

There was a bitter controversy between Mendelians and biometricians (Fisher Box, 1978, Gower, 1991). Pearson (1903) rejected Mendelism; he was convinced that most of the known human characters do not exhibit discrete types like Mendel's peas, so his approach was a biometrical one, dealing with the distributions of continuous variation. He showed (Pearson, 1903) that the theoretical expectations of Mendelism were not realized in his analysis of biometrical data. He assumed

- equally important Mendelian factors;
- the dominant and recessive phases being present in equal numbers;
- the different factors combining their effects by simple addition.

However, he found that the expected correlations between parent and offspring (0.33) worked out uniformly too low.

Fisher (1922) used both discrete Mendelian genetics and continuous biometrics to explain continuous variation. He showed (Fisher, 1918a,b; 1922; Fisher Box, 1978) how assortative mating could increase genetic correlations that Pearson, assuming complete dominance, equally important Mendelian factors, and equal frequencies of the alternative types of genes, had found to be too low.

Fisher (1922) showed that, under uniform genotypic selection balanced by occasional mutations, the large effect for parent and offspring due to dominance, accounted for about 32% of the total variance. This proportion depends on two ratios: the (unknown) gene frequencies for the pair of factors involved, and the (unknown) amount of dominance. As these two ratios vary, the proportion of dominance can take any value between zero and unity. He showed that dominance contributes nearly one third of the total variance in human measurements if the variance of the gene frequencies approaches zero; in this case the dominance ratio is independent of intensity of natural selection. The parental correlations are raised to values about 0.5, *not* by abandonment of Pearson's assumption of complete dominance as suggested by Yule (Fisher, 1918b, p. 428), but by assuming *differential importance* of the Mendelian factors in the first place, and different *proportions* in which they are present in the second place. With assortative mating all these proportions are modified (o.c., p. 410). Thus, Fisher agreed completely with Pearson (1903) in the size of the parental correlation, i.e., a dominance ratio* of 0.33 obtained under the conditions of complete dominance and random mating, and showed that this correlation could be raised to the

---

* The *dominance ratio* is the genetic correlation $r$ (e.g., the parental correlation) diminished by a fraction due to dominance effects. The effect of dominance is to reduce certain relationship correlations in the ratio $1 - \varepsilon^2/\sigma^2 = \tau^2/\sigma^2$, thus $r \bullet (1 - \varepsilon^2/\sigma^2) = 0.5 \bullet \tau^2/\sigma^2$, due to the relative contributions of the different phases of Mendelian factors and the relative frequencies with which they occur. Here, $\sigma^2$ is the total variance due to summation of a number of such factors associated at random, $\tau^2$ is the variance explained by genetic factors (the additive part), and $\varepsilon^2$ is the unexplained variance (the residue).

commonly found values of 0.50 by the phenomenon of assortative mating, thereby leaving virtually no room (about 5%) for the influence of environmental factors. To quote Fisher (1918a, p.218, [.] added):

> "This correlation [between father and mother] has the most profound effects upon the distribution of factors in the population, for, by virtue of it, similar phases of different factors become associated; the taller of two alternatives occurs in the same individual with the taller of another pair more frequently than random association would allow. Consequently, the variance is increased, the extremes of high and low stature becoming more common. With stature I find the variance increased from 5.611 sq. inches to 6.760 sq. inches, so that 17% of the total variance is due to mating of like with like."

Thus, assortative mating results in an *increase* in offspring variance.

Fisher (1918a; 1925) showed how the *intra-class* correlation coefficient could be used to partition the total variation into *non-heritable* fractions and *heritable* fractions, which itself can be partitioned into:

- additive gene action,
- dominance,
- gene interaction,

and showed that the excess of *sib* correlation over that found between parent and offspring, is due to innate and heritable causes (Fisher, 1918a). To analyze human variability, Fisher chose the squared standard deviation, the variance, because of its additive property: the variances contributed by independent causes sum to the total variance in the population. Many phenomena, previously expressed in terms of correlation, can be more appropriately thought of in terms of additive factors. The proper correlation coefficient in this framework is the *intra-class correlation coefficient*, which is formulated in terms of variance components (Fisher, 1925) and is almost identical to the

- strength of association index, $\omega^2$;
- correlation ratio, $\eta^2$;
- multiple correlation coefficient, $R_{y \cdot x_1 x_2 .. x_n}$;

developed for and used in different contexts (see Hays, 1988). Fisher (1928) showed that the distributions of all these coefficients are subsumed in the distribution of the multiple correlation coefficient, $R_{y.x}$ , used in the general linear model (see McCullagh and Nelder, 1989).

## 3 Overlooking Nurture: An Additive Effect and a Longitudinal Effect

As was shown in section 1, by using correlational analysis or covariance structure analysis, two distinct effects are disregarded: an *additive effect,* a difference in means, and a *longitudinal effect.* An example of the additive effect is the increase of 20 points in mean IQ on adoption as reported in the Skodak-Skeels (1949) study (cf. Freedman et al., 1978, Ch. 9; Jensen, 1973, p. 241; Loehlin et al., 1975, p. 297ff. Also, see Skeels, 1966, and Skeels and Dye, 1939). The point raised in this study is the following: Is intelligence determined by heredity, by environment, or by both? This Nature-Nurture controversy has been raging for more than three centuries, since John Locke (1632-1704) published his Tabula Rasa theory, according to which the mind is a blank tablet at birth, to be written on by circumstance. Attempts have been made to investigate this problem in a quantitative way. One of the best-known studies is about adoption and was conducted by Skodak and Skeels (1949) at an Iowa adoption agency. In the 1930s, they undertook a longitudinal study of children placed by this agency, from adoption (at the age of about six months) until adolescence. For each child, they measured the IQ at ages 2, 4, 7, and 14 years. They also measured each child's natural mother's IQ and the child's foster mother's educational level at the time of placement. They reported Pearson product moment correlations between each child's IQ and his natural mother's IQ, and between each child's IQ and his foster mother's educational level for the different time points. The results are as follows. The correlation between adopted children's IQs and their natural mothers' IQs first rises sharply, then stabilizes around a value of slightly larger than 0.30. The same pattern evolves for childrens' IQs and natural mothers' IQs when the children were reared by their natural parents (Honzik, 1957).

There was absolutely no contact between an adopted child and his natural mother after adoption, so the increase in the correlation over the years *cannot* be explained as a result of Nurture. On the whole, the correlation between the IQs of adopted children and the educational levels of their foster mothers are extremely low: they did not exceed a value of 0.10. Comparable results are reported by Jensen (1981, p. 102) for the Texas Adoption Study (Willerman, 1979): correlations of 0.32 between the IQs of children and their natural mothers' IQs, and 0.15 with their foster mothers' IQs. (Empirical correlations for various degrees of kinship are given by Jensen (1981, p. 92): 0.52 for parent-child, 0.19 for foster parent - child, and 0.09 for unrelated people.)

Therefore, this seems to be a strong piece of evidence for a hereditary theory of intelligence. However, the average IQ of the natural mothers was 86, but the average IQ of adopted children was 106. The improvement of 20 points in IQ can be largely attributed to the superior environment provided by foster parents. They were above average in economical security, educational and cultural status. The correlation coefficient does not pick up this change because it happens across the board: to each child's IQ about 20 points is added. Adding a constant amount to each observation is a shift of scale, which does *not* affect the correlation coefficient. Therefore, if inference is based only on correlation, one may arrive at a wrong or incomplete conclusion. In this case, the Nurture effect is completely overlooked. Comparable studies are reported by Flynn (1984), Jensen (1973, Ch. 7; 1981), and Loehlin et al. (1975).

## 4 Overlooking Nurture: Fathers and Sons

The second effect that may be disregarded by using correlations or covariance analysis is the *longitudinal effect*, which is hypothesized to be a continuing increase in the mean level of education in the population. This has been thoroughly discussed in: Van den Berg: 'The Netherlands are Becoming Always Smarter; Wait a Bit and Everyone Passes All Items at the Military Service Entrance Examination' (1989)*. In this study, a sample of 2854 Father and Son pairs were tested using the Raven Progressive Matrices Test as part of the Military Service

* In: Dossier Intelligentie I, Het Schoolblad, 8, pp. 5-8, 1989, The Netherlands.

**Table 1.** *Descriptive statistics for most important Father-Son variables with respect to Raven Test*

| Variable | mean | median | variance | stand. error | minimum | maximum |
|---|---|---|---|---|---|---|
| FOccup | 3.81 | 3.00 | 3.40 | 0.04 | 1.00 | 6.00 |
| FEduc | 68.22 | 71.00 | 124.41 | 0.21 | 16.00 | 84.00 |
| SEduc | 56.85 | 54.00 | 175.88 | 0.25 | 34.00 | 74.00 |
| FRaven | 21.61 | 22.00 | 45.53 | 0.13 | 0.00 | 40.00 |
| SRaven | 29.36 | 30.00 | 24.16 | 0.09 | 3.00 | 40.00 |
| DRaven | 7.75 | 8.00 | 49.78 | 0.13 | -22.00 | 33.00 |
| FBirth | 1936.39 | 1936 | 10.38 | 0.06 | 1931 | 1948 |
| SBirth | 1964.11 | 1964 | 0.11 | 0.01 | 1963 | 1967 |

**Table 2.** *Analysis of Variance for the Father-Son data with respect to Raven Test*

| Source of Variation | SS | DF | MS | F | Signif. of F |
|---|---|---|---|---|---|
| Within cells | 66326.05 | 2844 | 23.32 | | |
| FRaven | 75681.79 | 9 | 8409.09 | 360.57 | .000 |
| Total | 142007.84 | 2853 | | | |

**Table 3.** *Correlations between most important Father-Son variables*

| | FOccup | FEduc | SEduc | FRaven | SRaven |
|---|---|---|---|---|---|
| FEduc | 0.49 | - | | | |
| SEduc | 0.34 | 0.42 | - | | |
| FRaven | 0.37 | 0.50 | 0.32 | - | |
| SRaven | 0.15 | 0.23 | 0.48 | 0.30 | - |
| DRaven | 0.25 | 0.32 | -0.02 | -0.75 | 0.41 |

entrance examination in the period 1952 to 1982. The Fathers were born 1931 to 1948, and tested at age 18-19; the Sons were born 1963 to 1967, and tested at the same age.

### *The Additive Effect*

The main results for the Raven Test are presented in Table 1: an average Raven score of 21.6 for the Fathers ('FRaven') and an average score of 29.4 for the Sons ('SRaven'). The average *difference* between the Raven scores of Fathers and their Sons ('DRaven') is thus 7.8 points. The distribution of SRaven is shifted to the right compared to that of FRaven by 7.8 points (that is, by more than one standard deviation) on the Raven scale. The distributions for FRaven and SRaven are given in Figure 1. To test the additive effect, we use analysis of variance (or, equivalently, the $t$-test for differences). The average difference in Raven scores, DRaven, is 7.8, the standard error of the mean difference is 0.13, and the corresponding $t$-value is 59.62, which is significant at any level. Therefore, the mean of the distribution of SRaven is significantly different from that of FRaven. Figure 2 shows that DRaven, the *increase* in Raven scores from Father to Son when

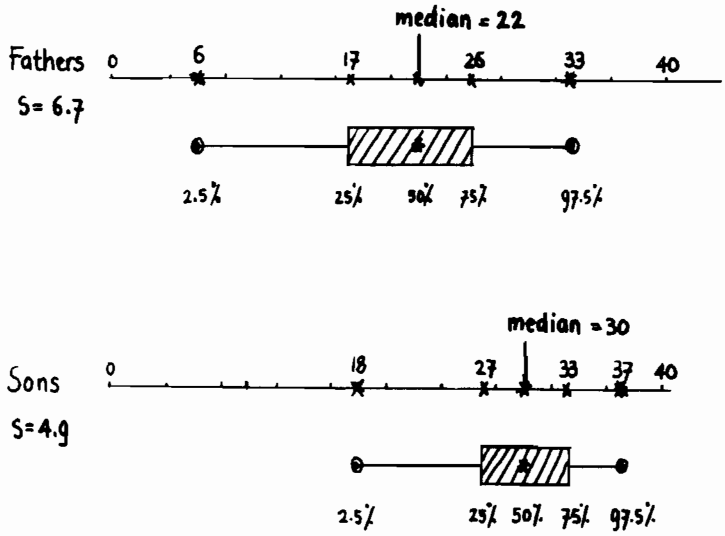


**Figure 1.** *Distribution of scores on the Raven Progressive Matrices Test for Fathers and Sons.*

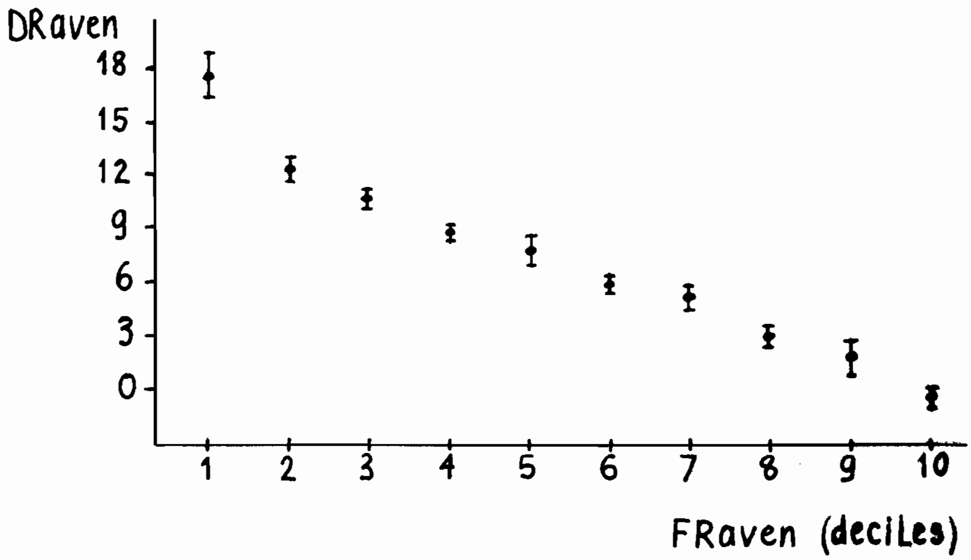


**Figure 2.** *Mean difference score DRaven and 95% confidence interval for the mean difference on the Raven Progressive Matrices Test given FRaven categorized into decile-intervals.*

conditioning on FRaven, is largest for Sons of low-Raven Fathers and that the increase diminishes linearly with increasing Father's Raven score, thereby *reducing* the dispersion of Sons' Raven scores as compared to their Fathers' Raven scores. From Figure 1, it can be read that the dispersion of SRaven (4.9) is significantly smaller than that of FRaven (6.7): $F = 1.87$ ($p < 0.01$). This finding points to absence of assortative mating for the factor of abstract intelligence (see quotation Fisher in section 2).

After categorizing FRaven by deciles (0-10%, 10%-20%,..., 90%-100%), we performed an analysis of variance. The results are shown in Table 2. The increase in Raven scores from Father to Son is very large: $F \sim 361$, which is significant at any level. Mean values for DRaven decrease monotonically from 17.79 (lowest decile) to -0.434 (highest decile). Hence, there is a tremendous difference, an *interaction effect*, which may be attributed to enriched education for everyone, and a more stimulating cultural and intellectual climate. However, on calculating the correlation between FRaven and SRaven, we find $r_{xy} = 0.30$, corresponding to the result reported in the Skodak-Skeels study (about 0.33), the Texas Adoption study (0.32), and by Honzik (1957) between IQs of natural mothers and their children (adopted or biological; see section 3). This *Nature* effect is discussed in greater depth in section 6.

*The Longitudinal Effect*

We also expected to find a longitudinal effect, a *change* in average Father-Son differences over the years, a random effect tied to the year in which the Father was born ('FBirth'). To test the longitudinal effect, we used variance components analysis (the multi-level model; see Aitkin and Longford, 1986; Goldstein, 1987; Longford, 1990; Prosser et al., 1991).With variance components analysis, the fixed effect is allowed to vary within a second factor. The model equation with only the fixed effect for DRaven denotes the 'first level', the equation with a random parameter for DRaven and a fixed parameter for FBirth denotes the 'second level'. Thus, at the second level, mean values of DRaven were allowed to vary within Year of Father's Birth. Thus, the average Father-Son difference on the Raven Test (DRaven) was assumed to vary randomly over the years. Both the fixed and the random effect for DRaven were tested using Longford's (1990) VARCL-program and Goldstein's ML3-program (Prosser et al., 1991). The random effect clearly was not significant. Therefore, it was concluded that average DRaven was constant over the years. The *fixed* main effect is significant, but too small to be of any importance. This may be due to two causes, first, the relatively small variance in FBirth, and, secondly, the very restricted range in Year of Son Birth ('SBirth'; 90% of the Sons are born 1964 or 1965, see Table 1). To be able to test the longitudinal effect, both FBirth and SBirth should have a larger variance. With respect to the longitudinal effect of Father-Son differences, further research is thus needed. However, a longitudinal effect for the Raven test for boys from schools of different intellectual level was shown by Meester and De Leeuw (1983, see below).

## 5 Further Evidence Concerning the Raven Test

As additional information about the Raven Progressive Matrices Test, we present the results of four cohorts (from 1952, 1962, 1972, and 1982) for Dutch secondary schools of varying level. Roughly speaking, four types of secondary schools can be distinguished according to intellectual level or demand. In order of decreasing intellectual level, these four types of secondary schools are: VWO (college), MULO (intermediate level secondary school), LBO (lowest level technical school), and LO (primary school only). For each secondary school type (VWO, MULO, LBO, and LO) and for each cohort, the percentage of boys of *high intelligence* is presented.

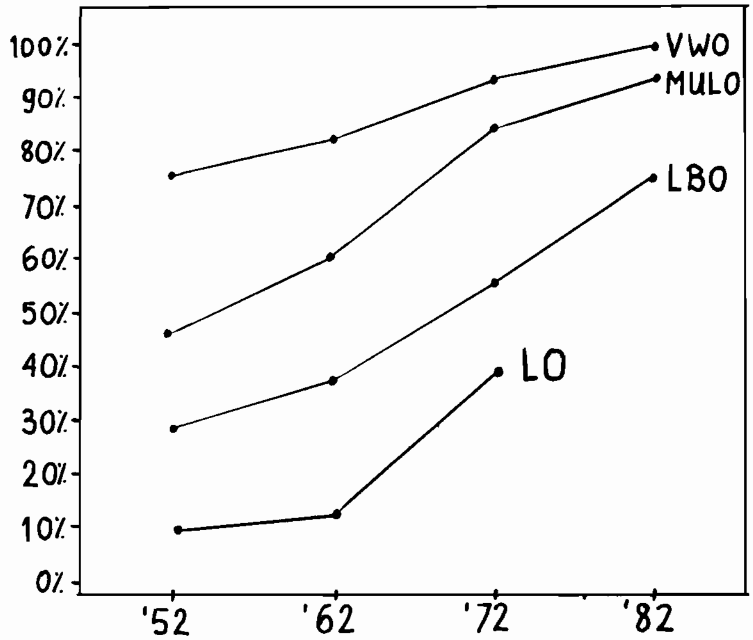


**Figure 3.** *Percentage of boys of high intelligence for schools of different intellectual level, tested using the Raven Progressive Matrices Test in 1952, 1962, 1972, and 1982.*

A person is designated of 'high intelligence' if his Raven score exceeds 24. The results are represented in Figure 3. From the Figure, it can be seen that the percentage of boys of high intelligence of VWO-level in 1952 is about the same as the percentage of boys of high intelligence of LBO-level in 1982. Thus, a Raven score higher than 24 was obtained by nearly as many working-class boys in 1982 as by higher-class boys in 1952. The increase is greatest for the lower-level schools, LBO and LO, the scores of MULO and VWO are converging. This is a clear indication of the presence of a longitudinal effect for the Raven test. In addition, increasingly more children choose higher levels of education: the entrance rates for lower-level secondary and technical schools are diminishing, and are growing for higher-level ones in the Netherlands. This was found by Tettero et al. (1978) for the period 1968-1977, and may imply, that, on average, working-class children have gained in IQ disproportionately as compared with higher-class boys, hence, they tend to choose higher educational levels. This confirms our earlier result concerning the *differences* between Fathers and Sons (DRaven) with respect to the Raven Test: Sons of low-Raven Fathers showed larger gains than Sons of high-Raven Fathers (see section 4).

*Intellectual Development: Food for the Brain*

The Raven Test measures abstract intelligence, well-known as 'transfer of training'. Improved education and cultural influences, such as information from newspapers, radio, television, movies, books, travelling abroad, and so forth, may be responsible for the increase in intelligence of the Dutch population as a whole. Analogous results for America are reported by Flynn (1984) and for Japan by Anderson (1982; see below).

Further hypotheses to explain the Raven effect were mentioned by Dr. J. Zaal from the Dutch National Institute for Psychological Testing (see Van den Berg, o.c.):

- *Increased motivation and ambition*: it is very important for people's self-esteem to be intelligent in our culture;

- *Increased test wisdom*: this effect may be important, but was rejected by Zaal as being the sole cause; to quote Zaal: 'No tomato-grower can improve his variety as quickly as the Dutch their people';

- *Inherited possibilities are more used in our culture*: abstract reasoning as measured by the Raven Test is more and more stressed. There is less pump and drill in school, less facts to learn but more understanding and stressing of relationships. The treatment of pupils is more democratic nowadays;

- *Spread breadthwise:* for all children in the Netherlands, language and mathematical achievement has remained approximately the same, even though the amount of time spent on these subjects has been diminished: the number of subjects in school has been enlarged. To quote Zaal again: 'Children achieve a lot more nowadays, but we are not aware of it because it is spread breadthwise'.

- *Need for intelligence*: Fathers nowadays transpose a quite different need for education and intelligence onto their Sons: a farmer in 1960 with low-level education and relatively low IQ has a Son with higher-level education and higher IQ, because this Son had to grow into a agrarian enterpriser, according to Vroon (University of Utrecht).

- *Intellectual potential and reaction range for IQ:* Jensen (1968, 1973, p. 175f.) proposed the hypothesis that the environment displays 'threshold' effects with respect to mental development. This means that environmental variations in one part of the total scale of environmental advantages have qualitatively or quantitatively different effects on the phenotype than in another part of the scale. Intelligence may show similar threshold effects. To quote Jensen (italics added):

> "In reviewing the literature reporting large shifts in IQ resulting from environmental changes, it was clear that all instances of larger gains in IQ were found in children whose environments had been changed from very poor to average or superior, while no IQ gains of comparable magnitude have been reported for children whose environments have changed from average to superior. This suggests a non-linear (or non-additive) effect of environment on mental development. Going from a typical slum environment to an average middle-class environment would presumably have a larger effect on IQ than going from a middle-class to an upper-class environment. We know that nutrition behaves more or less in this fashion with respect to stature. When the diet is deficient ( ), physical growth is stunted ( ). If we determined the hereditability of stature in a population that included a sizable percentage of persons whose nutrition had been inadequate for the full realization of their genetic potential for stature, we would find a *much lower heretability* than in a population in which everyone had adequate nutrition."

The mean gain in Raven scores for the Sons with respect to their Fathers is 7.8 points, and corresponds to a gain of 1.15 standard units in one generation. This gain is different for Sons of low-Raven and high-Raven Fathers (see section 4). The correlation between Fathers and Sons with respect to Raven is 0.30, falling short of every prediction of a hereditary factor that is one of the best measures for general intelligence, Spearman's $g$-factor . In fact, this value corresponds to Fisher's (1918b) prediction for a hereditary factor in the absence of assortative mating (a high percentage of error variance, see section 1). In the thirties, for *most* of the boys schooling was "deficient", thus their intellectual growth was "stunted" and, very likely, their "genetic potential for intelligence" was not fully realized (to keep up with Jensen's terminology).

Flynn (1984) reports an American IQ gain of about 14 points over 46 years (1932 - 1978), corresponding to about one standard unit. To quote Flynn (1984, p.74):

> "As for the total of environmental factors active at a particular time, the latest estimates attribute a maximum of .25 of IQ variance to between-families factors (...). Assuming that systematic environmental effects have a roughly normal distribution, one standard deviation of between-families difference in environmental quality makes a difference of 7.5 IQ points (the square root of .25 x 15 as value for SD for whites). Therefore, if we tried to explain our between-generations IQ gap in terms of within-environmental factors, we would have to say that the average environment in 1932 was 1.84 standard deviation (13.8 / 7.5) below average in 1978. This would put the Americans of 1932 at the 3rd percentile of environmental quality as measured today, which again taxes our credence."

Flynn mentions two possible causes of IQ gains over time. The first is increased test sophistication . According to Flynn, if increased test sophistication could explain IQ gains in isolation, it does not explain the puzzle posed by the combination of IQ gains and SAT-V losses in America. Probably, these SAT-V decline may be explained by Jensen's concept of 'realisation of intellectual potential': verbal intelligence may have been realized more fully with restricted schooling than *abstract* intelligence, thanks to the home-environment, availability of books, etc. The SAT-V effect may simply be one of 'decreasing marginal return'. In contrast, training in abstract intelligence in the home environment may have been little for most of the children, giving rise to large increases in IQ with increased schooling.

The last conclusion is in line with the second explanation of IQ gains over time: they are the benefits of a steady rising level of educational achievement. Flynn reports a study of Tuddenham (1948) about army data in the period 1918 - 1943. By weighting the 1918 test performance in terms of years of school completed, so as to match the 1943 educational distribution, explained about 55% of the mental test gains. But, as Flynn remarks, better education may be confounded with IQ: an educational elite will be to some degree a genetic elite as well, and the influence of superior genes for IQ will be confounded with better education. In our case, the steepest rise in Raven scores was for Sons of low-Raven Fathers, thus lending as much support for a craving for intelligence in intellectual slums as for an intellectual elite always consuming the best of education, thereby proving its superiority. What this study may show is that people outside of the intellectual elite do catch up lost ground, they *do* diminish differences in Raven scores if environmental conditions (schooling, quality of life, cultural level of a society) are available for everyone and are compulsory. Children must go to school, hence cannot have economical value for their elders until the age of 18. Anderson (1982) reports even stronger results for Japan: a IQ gain of 7 points over 23 years on the Wechsler test (see Anderson, 1982).

Some IQ's may be different from each other: selection and assortative mating may involve *general* intelligence; a sharp rise and then a decline in SAT-V scores need not be in contradiction with steadily increasing IQ scores in general: at some point in the evolution of a culture there may be a stronger need to enhance one's abstract intelligence than it may be necessary to enlarge or maintain a relatively high level of verbal intelligence.

*Mothers' IQ*

For the Father-Son data, also mothers' educational level was recorded. However, for the most part, the mothers had only the lowest level of education (primary school), so the variability of this variable was very low. Therefore, this variable was not included in the analysis. We may note that there are some studies reporting the influence of a sex-linked gene on spatial ability (e.g., Willoughby, 1927, Carter, 1932). The expected pattern of parent-child correlations for sex-linked characters is: $r_{FS} < r_{MD} < r_{MS} = r_{FD}$, where F, M, S, and D are father, mother, son, and daughter, respectively. Thus, for spatial ability, hereditability is highest for mothers and sons and for fathers and daughters. Later research (e.g., DeFries et al., 1979, Loehlin et al., 1978) questioned such a sex-linked hypothesis for intelligence.

## 6 Previous Analysis of the Father-Son Data

In Vroon, De Leeuw, and Meester (1986), quite different results for the above Father-Son data with respect to the Raven Test are reported. Vroon et al. deliberately disregard the large additive effect, the average difference of about eight points between Fathers' and Sons' Raven scores. Instead of this, their analysis is based on covariance analysis: they present a best fitting path model for the correlations between the variables measured. Their correlations are presented in Table 3. The best fitting path model is given in Figure 4.

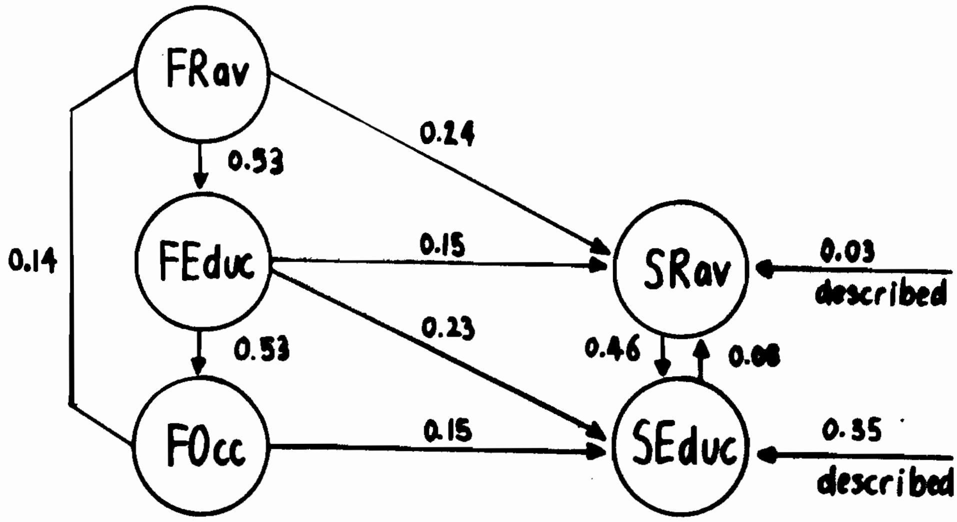


**Figure 4.** *Best fitting path model for the Father-Son-data with respect to theRaven Progressive Matrices Test as presented by Vroon c.s.*

The path model shows that the correlations within the set of the Father-variables are reasonably large, as they are within the set of Son-variables, however, the correlations between the two sets of variables are pretty low. From this, they conclude that there is not much evidence to defend either an environmentalist's position or a hereditarist's position. According to our exposition in section 2, this may be the result of the method used, namely, covariance structure analysis. Our stand is, that if they had used analysis of variance or any other method to estimate additive effects, they would have defended the environmentalist's position because of the large additive Nurture effect.

*Nature Effect*

The correlation between FRaven and SRaven is 0.30 and corresponds quite closely to the value of about one third for parent and offspring as predicted by Fisher in the case of absence of assortative mating (the dominance ratio, see section 2), as was the case for the correlations between the IQs of (adopted or biological) children and their natural mothers' IQs (Skodak and Skeels, 1949; Honzik, 1959; Skeels, 1966; Skeels and Dye, 1939; Willerman, the Texas Adoption Study, 1979; Jensen, 1981). Other correlations in Table 4 might indicate that Raven score on the one hand and occupation and educational level on the other, did not have a strong relationship in the socio-economical circumstances 30-40 years ago. Education and occupation were mainly working-class and executional level, less demanding abstract reasoning as measured by the Raven Test. Then this would imply that sexual selection did not concern abstract intelligence in the thirties.

Hence, our conclusion is that first, there is *both* a Nature *and* a Nurture effect, and secondly, that methodology has an important say in the Nature-Nurture controversy, almost certainly, in the effect of using an *inappropriate* methodology.

## 7 Discussion

Above, we presented the history of the Fisher-Pearson controversy, which results in distinct methodologies for giving a causal explanation of parent-offspring relationships. Under random mating, correlations between hereditary parent-offspring variables have an expected value of about one third. In the case of random mating, Pearson and Fisher's expected correlations are thus the same, about one third: Pearson's (1903) and Fisher's (1918b, 1922) dominance ratio. Assortative mating may increase parent-offspring relations up to about one half.

For the Raven Progressive Matrices Test, we found a large additive effect, a difference of about eight points between the score of Fathers and their Sons. We argued that this difference may be caused by environmental (cultural and educational) effects. Comparable results for America were reported by Flynn (1984), for Japan by Anderson (1982). Comparable results on adoption studies are reported by Skodak and Skeels (1949), Skeels (1966), Skeels and Dye (1939), and Willerman (1979); also, see Jensen (1973, 1981) and Loehlin et al. (1975). A longitudinal effect, a prolongued increase in Raven scores over the years, was hypothesized too for the Father-Son data, but could not be confirmed, probably due to the small variance in Year of Father Birth and Year of Son Birth. However, a longitudinal effect *was* shown to exist for mean Raven scores of boys from schools of different intellectual level in the period 1952 - 1982 (Meester and De Leeuw, 1983): a Raven score higher than 24 (signifying high intelligence) was obtained by nearly as many boys from lowest-level schools in 1982 as by boys from highest-level schools in 1952.

The increase in abstract resoning capacities from Father to Son is not evenly distributed given Father's score on this trait. Sons of low-Raven Fathers profit disproportionally as compared to Sons of high-Raven Fathers, thereby *reducing* the variance in Sons' Raven-scores to a considerable amount. This was confirmed in the Tettero et al. (1978) study for the period 1968 - 1977: over the years, children on the whole tended to choose schools of higher intellectual level. If there had been sexual selection (assortative mating) with respect to abstract intelligence, we should have found an *increase* in the variance of Sons' Raven scores (see quotation Fisher, section 2). In addition, the correlation between Father's and Son's Raven scores corresponds quite closely to the value predicted by Fisher for parent-offspring variables in the absence of assortative mating. So, we may conclude that environmental conditions and educational facilities favoured the development of abstract reasoning capacities: a strong *Nurture* effect has been present. The *Nature* effect is reflected in the parent-offspring correlations for IQ of about one third, for the Raven test in this study as well as for IQ in adoption studies ones (Skodak and Skeels, 1949, Honzik, 1957; Jensen, 1981, Ch. 7).

It was argued that a correlational study, comprising the calculation of Pearson product-moment correlation coefficients, or covariance structure analysis, cannot in general reveal the existence of additive effects. A Fisherian approach, using analysis of variance to estimate fixed effects or variance components analysis to estimate differential effects over time, is suited to reveal the existence of additive effects. This methodological difference was illustrated in the Skodak-Skeels study on adopted children and in the Vroon-De Leeuw-Meester data on Father-Son differences with respect to the Raven Test.

Vroon et al., in their Pearsonian approach, conclude that there is *neither* a Nature *nor* a Nurture effect. In our approach, the Nature effect is in accordance with Fisher's prediction: a correlation between FRaven and SRaven of nearly one third corresponding to the expected value for hereditary factors under the condition of random mating, and the Nurture effect shows up in increased intelligence during a period of increased technological advancements and demands, freedom from hard labour but compulsory and prolongued education for the youth, and a stimulating cultural environment. Therefore, we conclude that there is *both* a Nature *and* a Nurture effect. This difference in conclusions is attributed to the *methodologies* used to explore the relative contributions of Nature and Nurture. Clearly, if class-means are ignored, a sophisticated analysis of correlations is worthless. The Vroon et al. methodology is based on assumptions that are patently invalid and hence any conclusion so derived is worthless.

## References


Aitkin, M., and Longford, N.T. (1986). Statistical modeling issues in school effectiveness studies. *JRSS*, A, *149*, 1-43.

Anderson, A.M. (1982). The great Japanese IQ increase. *Nature, 297*, 180-181.

Carter, H.D. (1932). Family resemblances in verbal and numerical abilities. *Genet. Psychol. Monogr., 12*, 3-10.

DeFries, J.C, Johnson, R.C, Kuse, A.R, McClearn, G.E., Polovina, J., Vandenberg, S.G., and Wilson, J.R (1979). Familial resemblance for specific cognitive abilities. *Behav. Genet., 9*, 23-43.

Fisher, R.A. (1918a). The causes of human variability. *Eugen. Rev., 10*, 213-220.

Fisher, R.A. (1918b). The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Royal Soc. Edin.*, *52*, 399-433.

Fisher, R.A. (1922). On the dominance ratio. *Proc. Royal Soc. Edin.*, *42*, 221-241.

Fisher, R.A. (1928). The general sampling distribution of the multiple correlation coefficient. *Proc. Roy.Soc.* A, *121*, 654-673. Also collected in: R.A. Fisher: *Contributions to Mathematical Statistics.* London:Wiley.

Fisher, R.A.(1925; 1970, 14th Edition). *Statistical Methods for Research Workers.* Edinburgh: Oliver and Boyd.

Fisher, R.A. (1990). *Statistical Methods, Experimental Design, and Scientific Inference*: Re-issue by J.H. Bennett (Ed.) of: (i) Statistical Methods for Research Workers (1925), (ii) The Design of Experiments (1935), and (iii) Statistical Methods and Scientific Inference (1956). Oxford: Oxford University Press.

Fisher Box, J. (1978). *R. A. Fisher, The Life of a Scientist.* New York: Wiley.

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

Freedman, D., Pisani, R., and Purves, R. (1978). *Statistics.* New York: Norton.

Goldstein, H. (1987). *Multilevel Models in Education and Social Research.* Oxford: Oxford University Press.

Gower, J.C. (1991). Sir Ronald Aylmer Fisher, 1890 - 1962. *Math. Spectrum*, *23*, 76-86.

Hays, W.L. (1988, 4th Edition). *Statistics.* New York: HRW.

Honzik, M. P. (1957). Developmental studies of parent-child resemblance in intelligence. *Child Development*, *28*, 215-228.

Jensen, A.R. (1981). *Straight Talk about Mental Tests.* New York: Free Press.

Jensen, A.R. (1973). *Educability and Group Differences.* London: Methuen.

Loehlin, J.C., Lindzey, G., and Spuhler, J.N. (1975). *Race Differences in Intelligence.* San Francisco: Freeman.

Loehlin, J.C., Sharan, S., and Jacoby, R. (1978). In pursuit of the "spatial gene": A family study. *Behav. Genet.*, *8*, 27-41.

Longford, N.T. (1990). *VARCL: Software for variance component analysis of data with nested random effects (maximum likelihood).* User's Manual. Princeton, NJ: Educational Testing Service.

McCullagh, P. and Nelder, J.A. (1989, 2nd ed.). Generalized Linear Models. London: Chapman and Hall.

Meester, A.C., and De Leeuw, J. (1983). Intelligentie, sociaal milieu en de schoolloopbaan. Dept. Data Theory, University of Leiden, The Netherlands.

Pearson, K. (1903). On a generalised theory of alternative inheritance, with special reference to Mendel's laws. *Phil. Trans.*, *203* A, 53-87.

Prosser, R., Rasbash, J., and Goldstein, H. (1991). *ML3: Software for three-level Analysis.* User's Manual. Institute of Education, London: the University of London.

Skeels, H.M. (1966). Adult status of children with contrasting early life experiences: A follow-up study. *Child Dev. Mon.*, *31*, No. 3, Serial No. 105.

Skeels, H.M. and Dye, H.B. (1939). A study of the effects of differential stimulation on mentally retarded children. *Proc. Addr. Am. Ass. Ment. Deficienc.*, *44*, 114-136.

Skodak, M. and Skeels, H.M. (1949). A final follow-up study of one hundred adopted children. *J. Genetic Psych.*, *75*, 85-125.

Tettero, J.M., Bruens, J.F., and Brouwer, R.C.J.J. (1978). *De ontwikkeling van de deelname aan het voortgezet onderwijs, zowel landelijk als in onderwijskernen en omliggende gemeenten. Analyses en prognoses per schoolvorm.* Publikatie 15 van de Commissie Planprocedure, Ministerie van Onderwijs en Wetenschappen, 's-Gravenhage, The Netherlands.

Van den Berg, F. (1989a). Dossier Intelligentie I: Nederland wordt steeds slimmer; nog even en iedereen heeft alle vragen bij de dienstkeuring goed; *Het Schoolblad*, *8*, 5-8. (The Netherlands).

Van den Berg, F. (1989b). Dossier Intelligentie II: Hoe het onderwijs eerlijk werd, maar even onrechtvaardig bleef. *Het Schoolblad*, *9*, 34-37. (The Netherlands).

Vroon, P.A., De Leeuw, J., and Meester, A.C. (1986).Distributions of intelligence and educational level in fathers and sons. *Br. J. Psych.*, *77*, 137-142.

Willerman, L. (1979). Effects of families on intellectual development. *Am. Psych.*, *34*, 923-929.

Willoughby, R.R. (1927). Familial similarities in mental-test abilities. *Genet. Psychol. Monogr.*, *2*, 239-277.