Behavior Genetics, Vol. 20, No. 3, 1990

Differences in Heritability Across Groups Differing in Ability

Douglas K. Detterman,[1,3] Lee Anne Thompson,[1] and Robert Plomin[2]

Received 25 July 1989—Final 4 Dec. 1989

Differences in heritability and environmentality were assessed for 54 DZ and 86 MZ same-sex twin pairs between 6 and 12 years of age from the Western Reserve Twin Project. A principal-component score composed of the subtests of the WISC-R, PPVT, WRAT, and MAT represented each twin's cognitive ability. Using a modification of a regression technique developed by DeFries and Fulker (1985), it was possible to assess differential heritability and environmentality across ability level. A number of variants of this procedure were used and all yielded the same result: lower ability subjects show higher heritabilities and lower shared environmentality. This result is attributable to larger differences between DZ twins at low ability levels and to differences between MZ twins, which are either the same across ability level or are smaller at low ability levels. A possible explanation for this effect is a genotype–environment correlation in which higher-ability persons seek out better environments. The results from this study should be regarded as tentative but the methods used can be applied to other twins studies. Investigators should be aware of the importance of representing the low end of the distribution in their samples.

KEY WORDS: heritability; environmentality; intelligence; achievement; multiple regression.

Portions of this paper were presented at the Annual Meeting of the Behavior Genetics Association, June, 1989, Charlottesville, VA. Parts of this work were supported by Grants No. HD07176 and HD15516 from the National Institute of Child Health and Human Development, Office of Mental Retardation, the Air Force Office of Scientific Research and the Brooks Air Force Base Human Resources Laboratory, Project Lamp.

[1] Department of Psychology, Case Western Reserve University, Cleveland, Ohio 44106.

[2] Pennsylvania State University, University Park, Pennsylvania 16802.

[3] To whom correspondence should be addressed.

INTRODUCTION

It generally has been assumed that the structure of mental ability is uniform across the full range of ability, i.e, it has been assumed that correlations among mental tests were the same at the low end of the intellectual continuum as at the high end. A recent study (Detterman and Daniel, 1989) presented evidence, however, that suggests that the assumption of uniform correlations across the continuum may have been in error.

Detterman and Daniel found that correlations among basic tests of cognitive ability were higher for low-IQ subjects than for high-IQ subjects. They also analyzed the intercorrelations of the subtests of the WAIS-R and WISC-R using the standardization sample of each test. For both tests, intercorrelations among subtests were over twice as large for low-IQ groups as for high-IQ groups. In addition, for the five groups employed, there was a systematic trend for the average intercorrelation of the subtests to decline as the average IQ of the group increased. Further, these differences were found whether or not the appropriate corrections for restriction in range were applied. It would appear that the interrelationship of mental abilities differs across the ability continuum.

Detterman and Daniel's finding raises questions about what other characteristics may vary across the IQ continuum. Given the differences in correlations among subtests of standard intelligence tests, it does not seem implausible that heritability (h^2) or environmentality could differ for IQ and achievement tests across the ability continuum. The current study was aimed at investigating that possibility.

At least one study has found evidence suggesting that heritability varies across intellectual level. Reed and Rich (1982) reanalyzed data from the classic Reed and Reed (1965) study of mental retardation. They correlated midparent–offspring scores within three groups formed on the basis of midparent IQ. The lowest group had midparent IQs less than -2 standard deviations. The next group had midparent scores between -2 and $+1$ standard deviations, while the last group had scores $+1$ standard deviation above the mean. When offspring IQ was regressed on midparent IQ the last two groups were nearly equivalent but the lowest group had a regression value over three times as large as the other two groups.

Horn *et al.* (1979), in a reanalysis of the Texas Adoption Project, also found higher midparent–midoffspring regressions for a lower-IQ group split at the population mean, although the authors felt the difference was not a meaningful one because of small sample sizes. Vogler and DeFries (1983) reanalyzed data from the Hawaii Family Study of Cognition. None of the comparisons they evaluated were statistically significant but only four families fell below -2 standard deviations. When the sample was split into three groups, each having a proportion of subjects equal to the proportion of the Reed and Rich subdivisions, there was a statistically nonsignificant trend for high regression coefficients in the lower groups.

To our knowledge, there has not been an investigation assessing heritabilities in twins at low and high IQ levels. One reason may be that, until recently, there has not been a suitable methodology. Correlations, upon which the calculations of heritabilities are based, are highly influenced by restrictions in range that would be produced by dividing samples into groups. However, in a series of papers, DeFries and Fulker and their associates (DeFries, 1988; DeFries and Fulker, 1985; DeFries *et al.*, 1987; DeFries and LaBuda, 1989; LaBuda *et al.*, 1986; Zieleniewski *et al.*, 1987) have developed a simple yet elegant method for estimating heritability and environmentality using multiple regression techniques. As they point out, regression coefficients are not subject to the effects of restriction of range so the method can be used with groups having curtailed ranges. Although the technique can be used with any degree of relationship, it was initially developed with twin data and is easiest to conceptualize when twins are used.

In their original application of this methodology, the heritability of reading disorders was investigated. The sample consisted of members of twin pairs who had been identified as reading disabled. These twin pairs were matched with control twin pairs on the basis of sex, age, and zygosity. Data presented by LaBuda *et al.* (1986) illustrate the methodology in Fig. 1. Reading-disabled twins, the probands, have mean reading scores over a standard deviation from the population mean. The question is to what extent MZ and DZ cotwins of reading-disabled probands will be poor readers. If heredity does not contribute to the disorder, it would be expected that both MZ and DZ cotwins would regress to the mean of the population. That is, cotwins would show reading skills equivalent to those of the general population. On the other hand, if there was a heritable component to the disorder, MZ twins should show less regression to the mean than DZ twins. As shown in Fig. 1, reading-disabled MZ twins show less regression to the mean than DZ twins, suggesting a heritable component to reading disability.

The next question is whether reading disability is more heritable than reading skill, in general, or if good and poor readers demonstrate different heritabilities. A control group of subjects with normal or above-normal reading ability can be used to answer this question. Control subjects, also assessed on the same tests, can be tested for similar regression to the mean by identifying one of the twins (the better-reading twin, in this case) as the proband. If MZ and DZ control twins regress to a different degree than corresponding reading disabled twins, it can be concluded that heritability is different across groups. Thus, Fig. 1 not only shows heritability in each group as a function of the difference

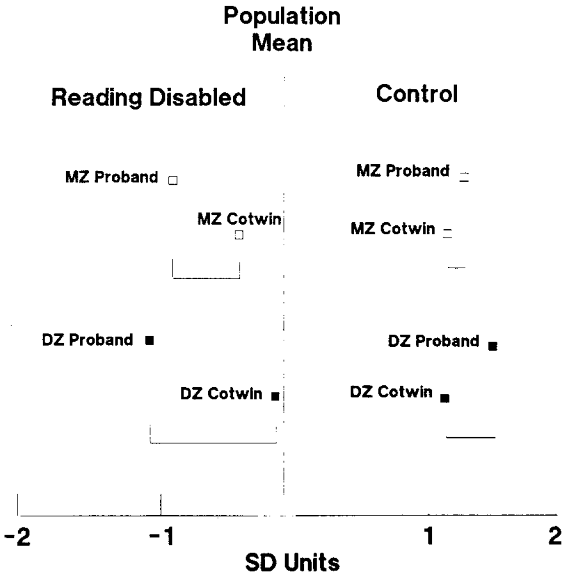

Fig. 1. The left panel shows differences between MZ and DZ reading-disabled probands and their cotwins. The right panel shows similar data for control subjects who are good readers. Data from LaBuda *et al.* (1986).

between MZ and DZ twins in the size of regression to the population mean of cotwins, but also allows a comparison of the differences in heritability by comparing the amount of this regression across groups.

Unfortunately, there are a number of factors that complicate the analysis just presented. MZ and DZ probands may have different means after selection. Proband and cotwin groups may differ in variance and covariance due to restriction in range produced by selection of probands to meet a criterion of cutoff. And finally, control and affected groups may differ in variance and covariance due to different selection criteria imposed on each group. Simply computing heritabilities and comparing them across groups would not work because heritabilities are correlational statistics which are highly affected by the factors mentioned previously, particularly restriction in range. Fortunately, the same information that could be obtained from computing heritabilities using correlations is available using multiple regression. The method developed by DeFries and Fulker (1985) allows estimation of group and individual heritability directly from data such as are presented in Fig. 1. More importantly, an extension of this method (LaBuda *et al.*, 1986) allows a direct test of differential heritability and environmentality across diagnostic group. The regression equation is

$$C = B_6P + B_7R + B_8D + B_9PR + B_{10}PD + B_{11}RD + B_{12}PRD + A,$$

where

C = cotwin score
P = proband score
R = coefficient of relationship (.5 for DZ and 1.0 for MZ)
D = diagnostic group (dummy coding of group membership)
A = intercept

Subscripted B's represent regression coefficients using the same numbering scheme that Fulker and DeFries and their associates have been careful to preserve across papers. B_6, the regression of proband score on cotwin score, is an estimate of the average environmentality (c^2) across groups. B_9, the regression of the Proband × Diagnostic-group interaction on Cotwin score, is an estimate of the average heritability (h^2) across groups. These estimates of c^2 and h^2 are only meaningful, of course, if c^2 and h^2 are the same across groups. To test for differential effects of c^2 and h^2, coefficients B_{10} and B_{12} are used. Each of these coefficients represents the interaction of diagnostic group with the term used to estimate the effect (P and PR). A statistically significant interaction indicates nonlinearity, i.e., the effect being tested (c^2 or h^2) differs across groups. Values for the above equation can be obtained from the multiple regression portion of any statistical package, which also will conveniently provide standard errors and a test of statistical significance for each of the regression coefficients.

When the above equation is applied to the data presented in Fig. 1, LaBuda *et al.* (1986) found that there was no evidence for either differential h^2 or c^2 across groups. A closer inspection of Fig. 1 will reveal that the ratio of MZ-to-DZ regression of cotwins from proband means is about the same for reading-disabled as for good readers. In the present paper, this same methodology is applied, in modified form, to assess the differential h^2 and c^2 across groups differing in mental ability as indexed by a composite of IQ and achievement tests.

METHOD

Subjects

Data for analysis were from the ongoing Western Reserve Twin Project. The current sample contains 54 pairs of same-sex DZ and 86 pairs of MZ twins. The number of male and female pairs is roughly equivalent for both MZ and DZ twins. All twins were between 6 and 12 years of age at time of testing. Mean age was 9.88 years (SD = 1.65) for DZ twins and 9.73 years (SD = 1.68) for MZ twins. Twins were recruited using state birth records and by nominations from schools. Zygosity was determined by a standard questionnaire assessing physical similarity which was administered by two separate raters. Doubtful cases on the zygosity rating scale were resolved by blood test.

The purpose of the Western Reserve Twin Project is to study intelligence, cognition, and school achievement with oversampling at the extremes of intelligence. The data for this study, however, were collected before any active effort was made to oversample. Mean WISC-R IQ was 104.2 (*SD* = 13.9) and 106.7 (*SD* = 14.11) for the DZ and MZ twins, respectively.

Procedures

All twins were given a battery of intelligence and achievement tests over three sessions of testing. The first session (1–1.5 h) was conducted in the twin's home. The second (2–4 h) and third (1–1.5 h) were conducted in the Psychology Department at Case Western Reserve University. In addition to the tests analyzed here, each twin was given two sets of tests of basic cognitive abilities and a series of measurements of basic physical characteristics.

Tests

Tests analyzed in the current study were the Wechsler Intelligence Scale for Children—Revised (WISC-R), of which all subtests but Mazes was administered, the Peabody Picture Vocabulary Test (PPVT), the Wide Range Achievement Test (WRAT), and the Metropolitan Achievement Test (MAT). Each of these tests, except the MAT, was individually administered to each twin by different examiners in different locations. The MAT was administered to groups of subjects, but in all cases twins were seated apart while taking the test. Scoring of each test was carefully checked by a testing assistant.

To obtain a composite measure, the age-standardized subtest scores from each of the tests were entered into a principal-component analysis and component scores for the first principal component were obtained for each subject. The first principal-component analysis was based on 1 score from the PPVT, 11 scores from the WISC-R subtests, 3 scores from the WRAT (Reading, Spelling, and Math), and 3 scores from the MAT (Reading, Math, and Language). Most subtests had a loading on the first principal component greater than .65. The only exceptions were five subtests from the WISC-R (loadings in parentheses): Digit Span (.52); Coding (.33), Object Assembly (.50), Block Design (.62), Picture Arrangement (.37), and Picture Completion (.44).

RESULTS

To determine if the same effects Detterman and Daniel (1990) found for the WISC-R standardization data hold for this sample, the subtest data from the

WISC-R for the twins in this study were analyzed using the same subgroups as in the Detterman and Daniel study. Twin pairs were treated as separate subjects. Each subject was placed into one of five groups based on their score on the vocabulary subtest. Correlation matrices were then computed for each subgroup. Correlations within each subgroup were corrected for restriction of range and the average correlation was then computed. Figure 2 shows the grouping used in IQ equivalent units and the average intercorrelation of the subtests. As in the Detterman and Daniel study, lower IQ groups showed higher intercorrelations among the subtests. The lower average correlations at the low end of the distribution for twins is probably due to the variability associated with the small sample size for that point. Whatever processes are operating to make correlations higher in the WISC-R standardization sample appear to be operating within this sample of twins.

Despite the problems with computing correlation coefficients in selected samples, as an initial exploration of the data, heritability was computed for the entire sample and for twin pairs above and below the group mean. Classification

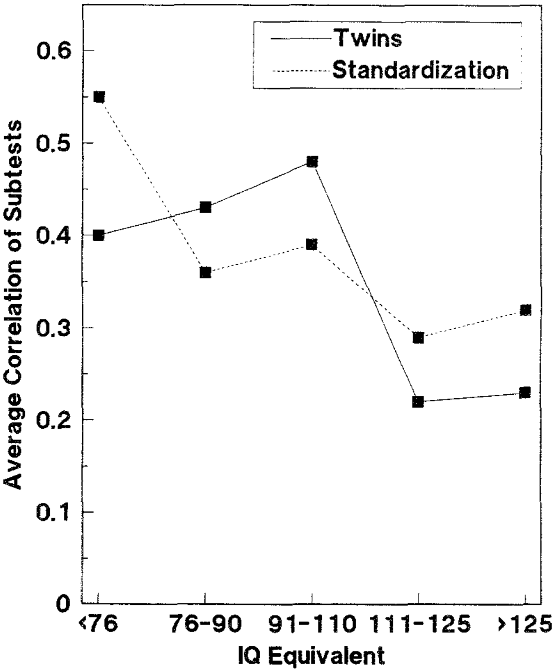

Fig. 2. Data from WISC-R standardization sample and WISC-R test results from twins of the Western Reserve Twin Project showing that subtests are more highly correlated at low IQ levels.

to group was made on the basis of the twin with the lowest component score. The group was split at the mean because that was the subdivision which would produce two groups having the same degree of restricted range. Intraclass correlation coefficients were computed and Table 1 shows the results for the whole group indicating h^2 to be about .49, an estimate consistent with past findings in the area. Based on intraclass correlation coefficients, it would appear that h^2 is higher in the low-IQ group. However, other subdivisions did not replicate this finding because of variance differences produced by restriction in range.

Another method of computing heritability more similar to the regression methods to be used was also conducted. One member of each twin pair was randomly selected as the proband. Based on the selected twins component score, the pair was assigned to either the low or high group as above. Standard bivariate regressions were then computed between the proband and the cotwin. Because selection of proband was random, this analysis was repeated 11 times and average correlations are shown in the bottom portion of Table 1 along with the heritability estimates computed from them. Results from this method are roughly equivalent to those from the method using intraclass correlations except that this method provides a lower estimate of heritability in the high group.

Division into Groups

The multiple regression methodology was extended according to the following logic: The reading study by LaBuda *et al.* (1986) can actually be thought of as comparing good and poor readers. This is not much different than comparing subjects high and low on cognitive ability. Further, there is no reason that this analysis should be confined to two groups. The method can be applied as easily to any number of groups.

Table I. Heritabilities Based on Intraclass Correlation Coefficients or Proband–Cotwin Regression Coefficients with Random Selection of Proband for First Unrotated Principal Component of IQ and Achievement Tests for Subsamples Divided at the Mean

Zygosity	All	Low	High
	Intraclass		
DZ	.60	.09	.32
MZ	.84	.59	.57
h^2	.49	1.00	.49
	Random selection—regression coefficients		
DZ		.22	.63
MZ		.71	.68
h^2		.98	.10

The same divisions shown in Fig. 2 were used to form twin groups on the basis of the lowest twin's component score. Proband score, P, was the lowest-scoring twin's score on the component and C, cotwin's score, was the other twin's component score. R was the coefficient of relationship, which was 1.00 for MZ twins and .50 for DZ twins. D, diagnostic-group membership, was dummy coded as -1, $-.5$, 0, $+.5$, and $+1$ for the lowest to the highest group, respectively. Interaction terms were represented by the products of the included main effect terms. Regression analysis was conducted using the SPSS Enter subcommand.

The effects of interest for the present purpose were the interactions, PD and PRD, which tested for differential c^2 and h^2, respectively. Both effects were highly significant. For PD, $t = 4.238, p < .001$, and for PRD, $t = -3.68, p < .001$, indicating that there were differential effects of c^2 and h^2 across groups. The negative t value indicates heritability is higher for lower IQ groups.

One difficulty with this approach is the way subjects were assigned to groups. Selecting the proband to be the lowest-scoring twin introduces bias. The mean of the probands for each group will be set by the selection criteria. The lowest group will have the lowest mean, etc. The cotwin mean must always be higher than the proband mean so differences between proband and cotwin score will be dependent on group membership. The lowest IQ group will show the largest difference between proband and cotwin. To determine if this bias had any effect on the multiple regression analysis, 101 Monte Carlo simulations were conducted. For each simulation, correlations among all variables were set to zero and random normal deviates were generated. The lowest score of a pair was designated the proband and the multiple regression analysis was conducted, as above, on the generated data. Averaged over all simulations, for PD (testing differential c^2)$t = .37. p = .51$, and for PRD (testing differential h^2)$t = .17$, $p = .53$. These results are what would be expected under the null hypothesis. Evidently, the selection procedure had no effect on the results of the regression analysis. In fact, the only effect statistically significantly different from zero in the simulation was the constant.

From the simulation data, it was possible to develop exact expectations for proband–cotwin differences for each ability level group under the null hypothesis of no correlation between twin pairs. That is, the simulation provided expected values that would be obtained from unrelated individuals randomly drawn from the population. This expectation could then be compared to obtained differences. To obtain the expectation, the cotwin–proband difference for each of the 101 simulations was computed for each ability level group and then averaged over simulations. Figure 3 shows the MZ and DZ cotwin–proband differences as well as the expected difference from the simulation. It appears that, as ability decreases, MZ cotwin–proband differences decline relative to expectation but DZ cotwin–proband differences clearly become larger as ability level declines. Al-

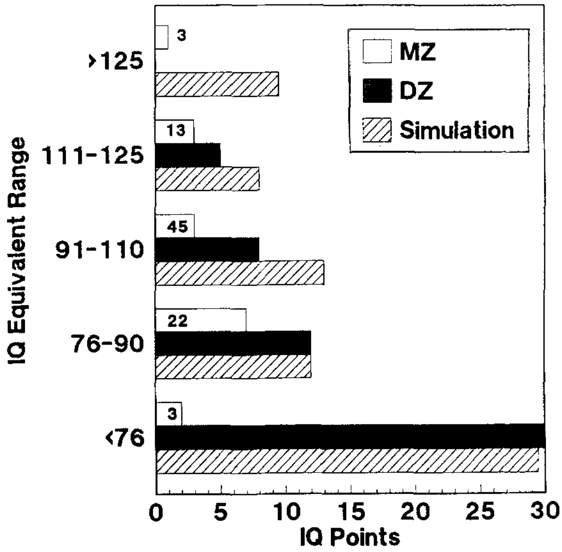

Fig. 3. Differences between cotwins and proband for different ability-level groups for MZ and DZ twins and for data obtained from 101 simulations. Simulation data show expected results if cotwin and proband scores are uncorrelated. Numbers at the base of MZ and DZ bars are the number of twin pairs in each group.

though the extreme groups show these differences most dramatically, the middle groups show the same effect. In the 111–125 group, MZ and DZ cotwin–proband differences are about half of expectation. However, in the 76–90 group DZ cotwin–proband differences meet expectations, while MZ cotwin–proband differences are about half of expectation as they were for the 111–125 group. In the lowest group, MZ twins reach only 10% of expectation, suggesting that MZ twins are getting more alike at the lowest level. However, the trend for DZ twins to become more dissimilar at lower ability levels is clearer than the trend for MZ twins to become more alike.

Heritability is inversely related to ability level, while environmentality is directly related to ability level. DZ twins are most similar when high in ability and least similar when low in ability. These differences between DZ twins differing in ability level produce the differences in heritability and environmentality.

Division into Groups—Midtwin Criterion

To investigate further the effect of group division on regression analysis, another method of assigning probands to groups was used. Instead of making assignment to groups on the basis of the lowest scoring twin's score, assignment was made based on the midtwin score. That is, the twins' component scores were averaged and assignment to groups was made on the basis of this averaged score, using the same cutoffs as shown in Fig. 2. The proband was still defined as the lowest-scoring twin of the pair. All other aspects of the regression analysis were identical to the previous analysis.

The effects obtained by this method were very nearly identical to the last. For PD, $t = 4.034, p < .001$, and for PRD, $t = -3.794, p < .001$, indicating differential effects of c^2 and h^2, respectively, across groups.

Assignment to group on the basis of midtwin score has its own potential bias. Twin pairs having the most discrepant scores will tend to be assigned to middle groups since, to be highly discrepant, a wider range of the ability continuum would be used. For example, the most discrepant scores, which would be the highest and lowest possible score, would average to the mean of the distribution. If MZ twins are more highly correlated with each other than DZ twins, then this bias would have the largest effect on DZ twins.

Division into Groups—Random Assignment of Probands

A method of assignment of subjects to groups that avoids the problems associated with the previous two methods is assignment to groups based on a random choice of one twin or the other as proband. This method has the disadvantage that there are many potential designations of proband and cotwin (2^n) so no single ordering can be counted on to reflect the data accurately. To avoid this problem, 10 analyses were conducted and the results averaged. For each analysis, one member of the pair was randomly designated as proband. Groups were then formed on the basis of proband score according to the divisions shown in Fig. 2.

The results of the 10 analyses were averaged. For the PD effect, averaged results were $t = 2.831, p = .029$, and 2 of 10 t values were not significant (largest $p = .108$). For the PRD effect, averaged results were $t = 2.791, p = .040$, and 3 of 10 t values were not statistically significant (largest $p = .15$). Even using this conservative method of assigning subjects to conditions, there is little question that both the PD and the PRD effects are statistically significant supporting the finding from other methods of differential c^2 and h^2 across groups.

Continuous Regression—Random Assignment of Probands

If it is possible to use five groups to do this regression analysis, it should be possible to replace group membership with a continuous variable. The continuous variable of ability is actually the proband's component score. Substituting P^2, the proband's component score squared, for D, diagnostic group, in the previous regression equation, yields the following:

$$C = B_6P + B_7R + B_8P^2 + B_9PR + B_{10}PP^2 + B_{11}RP^2 + B_{12}PRP^2 + A.$$

Although this equation looks odd, a little study will indicate that it is completely proper. B_8 formerly accounted for the variance from the mean of the diagnostic group but now each proband can be considered a diagnostic group so the B_8 term accounts for the quadratic regression effects of the probands score on the cotwins score. Similarly, in other terms of the equation P accounts for linear regression effects and P^2 accounts for interaction or nonlinear effects.

The above regression equation was computed using P^2 in place of D and designating the twin with the lowest component score the proband. In this analysis, both the PD effect ($t = -1.612, p = .11$) and the PRD effect ($t = 1.76$, $p = .08$) approached statistical significance but were not significant. However, when P^3 was substituted for P^2, both the PD ($t = 3.67, p < .001$) and the PRD effect ($t = -3.65, p < .001$) were highly significant. Since the component scores have a mean of zero, using P^2 makes all scores positive and eliminates the difference between positive and negative scores. P^3 preserves the sign and assesses differential h^2 and c^2 as a cubic function of the probands score. To ensure that this was the source of the problem, the analysis was repeated using P^2 in place of D but if P was less than zero, P^2 was multiplied by -1 to preserve the sign of the proband component score. In this analysis, both the PD ($t = 4.232, p < .001$) and the PRD ($t = -4.11, p < .001$) effects were statistically significant. Thus, when appropriately implemented, using individual scores in place of groups demonstrates the same effects.

A more direct method of simultaneously estimating c^2 and h^2 in the total sample and assessing differential h^2 and c^2 as a linear function of proband's score can be obtained from the following hierarchical regression analysis (We thank an anonymous reviewer for suggesting this analysis.):

$$C = B_6P + B_7R + B_8PR + B_9P^2 + B_{10}P^2R + A.$$

The first three terms are entered on the first step and the last two on the second step. B_6 and B_8 estimate c^2 and h^2 in the entire sample. B_9 and B_{10}, computed on the second step, estimate differential c^2 and h^2, respectively, as a linear function of proband's scores. This method has the advantage of producing estimates of c^2 and h^2 for the entire sample and also estimating differential h^2 and c^2.

This method was applied, as before, with random selection of probands. The analysis was repeated 10 times and average values are reported. h^2 was estimated to be .59 ($t = 2.69$, 6 of 10 t values, $p < .01$) and c^2 was .24 ($t =$ .98, 1 of 10 t values, $p < .01$). This analysis also provided strong support for differential h^2 ($t = -3.64$, 8 of 10 t values, $p < .01$) and differential c^2 ($t =$ 3.82, 8 of 10 t values, $p < .01$). This analysis is in good agreement with results obtained from other methods but may be preferred because it provides both h^2 and c^2 estimates as well as a test of differential h^2 and c^2 in a single analysis. In fact, it might be argued that this should be the standard regression method for computing estimates of heritability because it simultaneously indicates if the estimates of heritability and common environment can appropriately represent the entire group.

DISCUSSION

It would appear, from these data, that both h^2 and c^2 differ with IQ level. Lower-IQ groups appear to have higher heritability and lower shared environmentality. Why does the effect occur and why have other researchers had a difficult time finding it? The empirical reason for increased heritability at lower IQ levels is reasonably clear. DZ twins become increasingly dissimilar as ability level decreases and MZ twins may become more alike, or at least remain as similar, over ability level. What causes these differences is less clear.

Before discussing the potential causes of the effects found here, several cautions are in order. First, the number of subjects is relatively small, so even though the effects were generally highly statistically significant in a variety of analyses, the results should be considered tentative until replicated with larger sample sizes. Second, the results are importantly dependent on good representation at the low end of the distribution. A better test of changes in h^2 and c^2 over ability level could be made with oversampling at the extremes. That is what the Western Reserve Twin Project will be doing and so data should be available for a stronger test of this hypothesis. Finally, there are negative findings in the literature, and although each study showed trends suggesting the results found here, those negative results urge caution in the interpretation of the current results (see below for a discussion of potential reasons for differences in results). Taken together, these cautions suggest that the current results should be taken as tentative. However, the issue of differences in heritability is important enough that other investigators should be encouraged to examine their data for similar effects and, whenever possible, ensure that the low end of the distribution is adequately represented in their samples.

There are a number of potential explanations for the effects found here. Falconer (1981) describes conditions which can cause differences in heritability and environmentality across the range of a variable. In all cases, the differences are due to differences in skew of the environmental and genetic variance. This difference in skew can be due to scale effects or genotype–environment interaction. Scale effects are irregularities in measurement which can cause differences in h^2 and c^2 that often can be eliminated by transformation. Genotype–environment interactions can produce differences in one of two possible ways. First, persons who have good environment may show less genetic variation. For example, flower varieties might show less variation when all are grown in good soil than when grown in bad soil. Or milk cows may show less variation in milk production under good environmental conditions. Second, individuals high in a characteristic may be more affected by environmental variation than those low in the characteristic. For example, when high-producing milk cows bred in Wisconsin are moved to Texas, their milk production may be more affected by the environmental change than low-producing Wisconsin cows. In all of these cases, what changes across the continuum is the relative proportion of variance due to environmental and genetic causes. In all of the examples above, hereditary influences would be larger for persons low on the characteristic under consideration.

The effects observed in this study could have been produced by differences in assortative mating. If there had been lower levels of assortative mating for the parents of lower-IQ subjects, low-IQ DZ twins would be more dissimilar but the similarity of MZ twins would have been unaffected. There is no support for this explanation in our data, as neither years of education nor an index of SES shows consistent differences in assortative mating large enough to account for the differences in heritability across ability level.

Another possibility is that there are nonadditive major gene effects which depress IQ at the low end of the distribution. Such effects would make low-IQ DZ twins less similar but would not affect the similarity of MZ twins. Grayson (1989), extending the work of Eaves *et al.* (1978), has investigated the potential effect nonadditivity could have on estimates of heritability (since it is usually assumed to be zero) and concludes that it may as much as double them. Estimates of heritability calculated for this sample are roughly consistent with this possibility, although they are at the extreme of Grayson's estimates because heritability in the low-IQ group is at least twice that of the high-IQ group (and sometimes more).

As both Grayson and Eaves *et al.* point out, when the ratio of DZ-to-MZ covariances is less than .5, nonadditivity must be involved, and when the ratio is greater than .5, shared environmental effects are implicated. Inspection of Table I indicates that the ratio of correlations of DZ-to-MZ twins for the low-IQ group is less than .5, which supports the possibility of nonadditive effects. On the other hand, the ratio of the DZ-to-MZ correlations for the high IQ groups is equal to or greater than .5, suggesting the effects of shared environment. These ratios are consistent with the argument that at least a portion of the differences in heritability across IQ levels is due to nonadditive effects at the low end of the ability distribution.

How do these explanations, particularly those of genotype–environment interaction and correlation, relate to the current results? There are certainly many potential explanations. One possibility is that low-ability persons are less able to seek out or take advantage of good environments than are more capable persons. At the low end of the ability continuum, whether or not subjects found an environment most suited to their genotype would depend on chance factors. Those high in ability, on the other hand, would actively seek appropriate environments and subsequently be affected by them. Becuase genes and environment are the same for MZ twins, they would be equally similar at all ability levels. Each member of an MZ twin pair would show the same degree of genotype–environment coaction. But because DZ twins differ genetically, a DZ twin high in ability might seek out appropriate environments, thereby increasing his/her intelligence but the cotwin, low in ability, might show just the opposite effect and, instead of seeking out better environments, be less reactive to his poor environment. This would work to make DZ twins more dissimilar. At the theoretical extreme, one DZ twin could show maximal phenotypic gain from positive environmental effects, while the other could show a phenotypic handicap from the same environment.

Why have previous studies, with the exception of Reed and Rich, not found this effect that seems so large in this study? There are several possibilities, the most obvious of which is the number of subjects which are included in the low-IQ range. Low-IQ subjects require special efforts to recruit. They often are not sympathetic to research efforts. Both our study and the Reed and Rich study included representation at the low end of the scale where differences are largest, and this may explain why both of these studies found effects not found by the others. Vogler and DeFries (1983) suggest that inclusion of low-IQ subjects may represent an altogether different distribution known to exist at the low end of the ability continuum. But individuals from that portion of the distribution are usually unable to take the kinds of cognitive tests given here. None of the subjects included in this study reflect that degree of impairment. Another reason we may have obtained these results is that we used a particularly large battery of intelligence and achievement tests. These tests may better represent the domain of mental ability than the more limited sets of tests used by previous investigators.

The preceding comments are particularly relevant to the study recently reported by Capron and Duyme (1989). In that study, adoptees were carefully selected to form a completely crossed design between high- and low-SES and adoptive and biological parents. The study showed only main effects, without the interaction found in our study. Supporting the comments made by McGue (1989), an important difficulty with this study is that the lowest IQ in each of the four groups formed was 68, 91, 91, and 99. The lowest mean IQ for any group was 92.4. If our study had been confined to these IQ ranges, we would have essentially replicated the Capron and Dumye results.

Despite differences in results, there is accumulating evidence suggesting that heritability differs for low- and high-ability subjects. It is important to determine if this effect actually exists, and if it does, it will be crucial to determine it cause. If heritability is higher at the low end of the ability continuum, it could have important implications for theory and practice in education and psychology.

ACKNOWLEDGMENTS

We would like to thank D. W. Fulker, who suggested the random assignment analysis, and J. C. DeFries, who suggested the P^2 analysis.

REFERENCES

Capron, C., and Duyme, M. (1989). Assessment of effects of socio-economic status on IQ in a full cross-fostering study. *Nature* **340:**552–554.

DeFries, J. C. (1988). Colorado reading project: Longitudinal analysis. *Ann. Dyslexia* **38:**120–130.

DeFries, J. C., and Fulker, D. W. (1985). Multiple regression analysis of twin data. *Behav. Genet.* **15:**467–473.

DeFries, J. C., and LaBuda, M. C. (1989). Differential prognosis of reading-disabled children as a function of gender, socioeconomic status, IQ, and severity: A longitudinal study. *Read. Writ.* **1:**25–36.

DeFries, J. C., Fulker, D. W., and LaBuda, M. C. (1987). Evidence for a genetic aetiology in reading disability of twins. *Nature* **329:**537–539.

Detterman, D. K., and Daniel, M. H. (1989). Correlations of mental tests with each other and with cognitive variables are highest for low IQ groups. *Intelligence* **13:**349–359.

Eaves, L. J., Last, K. A., Young, P. A., and Martin, N. G. (1978). Model-fitting approaches to the analysis of human behavior. *Heredity* **41:**249–320.

Falconer, D. S. (1981). *Introduction to Quantitative Genetics* (2nd ed.), Longman, London.

Grayson, D. A. (1989). Twins reared together: Minimizing shared environmental effects. *Behav. Genet.* **19:**593–604.

Horn, J. M., Loehlin, J. C., and Willerman, L. (1979). Intellectual resemblance among adoptive and biological relatives: The Texas Adoption Project. *Behav. Genet.* **9:**177–207.

LaBuda, M. C., DeFries, J. C. and Fulker, D. W. (1986). Multiple regression analysis of twin data obtained from selected samples. *Genet. Epidemiol.* **3:**425–433.

McGue, M. (1989). Nature-nurture and intelligence. *Nature* **340:**507–508.

Reed, E. W., and Reed, S. C. (1965). *Mental Retardation: A Family Study*, Saunders, Philadelphia, PA.

Reed, S. C., and Rich, S. S. (1982). Parent-offspring correlations and regressions for IQ. *Behav. Genet.* **12:**535–542.

Vogler, G. P., and DeFries, J. C. (1983). Linearity of offspring-parent regression for general cognitive ability. *Behav. Genet.* **13:**355–360.

Zieleniewski, A. M., Fulker, D. W., DeFries, J. C., and LaBuda, M. C. (1987). Multiple regression analysis of twin and sibling data. *Personal. Indiv. Diff.* **8:**787–791.

Edited by D. W. Fulker