Research Article

SOCIOECONOMIC STATUS MODIFIES HERITABILITY OF IQ IN YOUNG CHILDREN

Eric Turkheimer, Andreana Haley, Mary Waldron, Brian D'Onofrio, and Irving I. Gottesman

University of Virginia

Abstract—*Scores on the Wechsler Intelligence Scale for Children were analyzed in a sample of 7-year-old twins from the National Collaborative Perinatal Project. A substantial proportion of the twins were raised in families living near or below the poverty level. Biometric analyses were conducted using models allowing for components attributable to the additive effects of genotype, shared environment, and nonshared environment to interact with socioeconomic status (SES) measured as a continuous variable. Results demonstrate that the proportions of IQ variance attributable to genes and environment vary nonlinearly with SES. The models suggest that in impoverished families, 60% of the variance in IQ is accounted for by the shared environment, and the contribution of genes is close to zero; in affluent families, the result is almost exactly the reverse.*

Although the heritability of cognitive ability in childhood is well established (McGue, Bouchard, Iacono, & Lykken, 1993; Plomin, 1999), the magnitude, mechanisms, and implications of the heritability of IQ remain unresolved. Historically, the most controversial question surrounding the heritability of intelligence is whether genetic effects on IQ place serious constraints on the effectiveness of efforts to raise IQ, either by improving impoverished socioeconomic conditions or by exposing children to remedial educational programs such as Headstart (Herrnstein & Murray, 1994; Jensen, 1981). Adding to the controversy is an apparent contradiction between studies using different methodologies to study the development of cognitive abilities: Studies of correlations among twins or adoptees and their biological and adoptive parents typically yield large genetic effects and relatively smaller effects of family environment, whereas studies that compare the mean IQs of children rescued from poverty with the IQs of their parents or impoverished siblings often find large differences that are attributed to the environment (Turkheimer, 1991).

One possible resolution of this paradox is that the effect of family environment on cognitive ability could be nonlinear (Jensen, 1981; Scarr, 1981). If differences between impoverished environments and adequate ones have large consequences for cognitive ability, but differences between adequate and enriched environments do not, one would expect amelioration of impoverished status to show a substantial effect, whereas correlational findings based on middle-class family members in typical twin and adoption studies would not. Unfortunately, genetically informed empirical tests of this hypothesis have been rare, for two reasons. First, few twin studies include children from highly impoverished backgrounds, and because impoverished parents are generally unable to adopt, impoverished environments are systematically censored in adoption studies (Stoolmiller, 1999). Second, until very recently, there have been severe restrictions on researchers' ability to model interactions between a contin-

uous variable like socioeconomic status (SES) and latent genetic and environmental influences on cognitive ability.

Nevertheless, several previous studies have addressed differential heritability as a function of race, social class, or parental education. Scarr-Salapatek (1971) obtained a sample of twins from the Philadelphia school system and used standardized test scores as a measure of ability. SES was estimated from census-tract information. A major limitation of the study was that it did not include zygosity information about the twin pairs; instead, analyses were based on comparisons of same-sex pairs (that combined monozygotic, or MZ, twins and dizygotic, or DZ, twins) and opposite-sex pairs (all DZ twins), a method with considerably less statistical power than the classical twin design (Eaves & Jinks, 1972). For both Black and White children, estimated heritabilities were lower in children from families in lower-SES census tracts than in those from middle- or higher-status tracts.

Scarr subsequently (1981, chap. II.4) obtained an independent sample of 160 Black and 212 White twin pairs from the Philadelphia school system. Unlike in the earlier study, SES was measured for each family, and zygosity was determined by blood typing. The twins were administered a battery of intelligence tests, including the Raven Standard Progressive Matrices, Peabody Picture Vocabulary Test, Columbia Mental Maturity Scale, and Benton Revised Visual Retention Test. MZ and DZ twin correlations were computed separately for Black and White twin pairs and for low- and high-SES groupings. Differences between MZ and DZ twin correlations were consistently higher for the White pairs; no consistent differences emerged when the twins were grouped by SES within race.

Fischbein (1980) studied a Swedish sample of 87 MZ and 126 DZ pairs of 12-year-old schoolchildren. SES was estimated from parental education and occupation and used to divide the twin sample into three groups. Ability scores were obtained from a self-administered test that included measures of verbal ability and inductive reasoning. For both measures, heritabilities were highest in the high-SES group and lowest in the low-SES group.

Rowe and his colleagues have undertaken several studies of differential heritability in large national data sets. Van den Oord and Rowe (1997) used hierarchical linear models to study differential genetic and environmental contributions to scores on the Peabody Individual Achievement Test in a sample of siblings, half-siblings, and cousins drawn from the National Longitudinal Survey of Youth. They reported only the scantiest evidence of covariation between the magnitude of genetic and environmental variance components and measures of the home environment in which the children were raised.

More recently, Rowe, Jacobson, and van den Oord (1999) analyzed a large sample of twins (176 MZ pairs; 347 DZ pairs), siblings (795 pairs), half-siblings (269 pairs), cousins (118 pairs), and unrelated siblings reared together (204 pairs) from the National Longitudinal Study of Adolescent Health. Parental education was employed as an indicator of SES, and ability was measured using an abridged version of the Peabody Picture Vocabulary Test–Revised. The authors used DF analysis (DeFries & Fulker, 1985; LaBuda & DeFries, 1990) to estimate the

 623

biometric parameters and their interaction with continuously measured education. Results showed that both heritability and shared environment interacted significantly with education: Heritability increased and effects of shared environment decreased as parental education increased. Similar findings using DF analysis have been reported by Thompson, Tiu, and Detterman (1999).

METHOD

In the current study, we used data from the National Collaborative Perinatal Project, which included a large national sample of American mothers, who were enrolled into the study during pregnancy ($n = 48,197$), and their children ($n = 59,397$), who were followed from birth until age 7 (Nichols & Chen, 1981). Participants were recruited from 12 urban hospitals around the country and included a high proportion of racial minorities and impoverished families. Extensive medical, psychological, and socioeconomic data were obtained for the mothers during pregnancy, and for the children at birth and at ages 8 months, 1 year, 4 years, and 7 years. The Wechsler Intelligence Scale for Children (WISC) was administered at age 7. Verbal IQ (VIQ), Performance IQ (PIQ), and Full-Scale IQ (FSIQ) scores were estimated from a slightly reduced set of WISC subtests. Socioeconomic scores were obtained at mother's registration in the study and at the 7-year evaluation. These scores were based on the 100-point system of Myrianthopoulos and French (1968) and computed from a linear combination of parental education, occupational status, and income.

The sample included 623 twin births. Of these, 320 pairs with complete data regarding IQ, SES, and zygosity remained at the 7-year follow-up. Twins remaining in the sample at 7 years of age did not differ from twins lost to the sample in terms of birth order, mother's marital status at birth, family SES at birth, race, or family income at birth. One additional DZ pair was eliminated as an outlier, because of an 81-point difference between the IQ scores of the twins; the twin with the lower IQ was identified as brain damaged at birth. Of the remaining 319 pairs, 114 were monozygotic and 205 were dizygotic. Of the DZ pairs, 81 were same-sex pairs and 124 were opposite-sex pairs. There were no significant mean differences for any analysis variables between same- and opposite-sex DZ pairs, and no differences in the twin correlations, so the opposite-sex pairs were combined with the same-sex pairs in all analyses.

The twins were classified as 43% White, 54% Black, and 3% "other." The sample included a high proportion of impoverished families. The median number of years of education of the head of household was between 10 and 11 years, and 25% of household heads were not educated past the ninth grade. The median occupation was "service worker"; 25% of the household heads received occupational ratings of "laborer" or lower, including 14% with no occupation. The median family income was between $6,000 and $7,000 annually, equivalent to $22,100 in 1997 dollars, the most recent year for which an equivalent scale is available. Twenty-five percent of the families had incomes below the 1973 poverty level for a family of four (U.S. Census Bureau, 2002).

In traditional biometric analyses of cognitive ability (Plomin, De-Fries, McClearn, & McGuffin, 2000), variation in IQ is partitioned into three independent components attributable to additive variation in genotype (A), shared environment (C), and nonshared environment (E). Our goal in the current analyses was to fit a model in which the magnitude of the variance components themselves varied as a continuous function of SES. That is, rather than modeling children's IQ as a simple
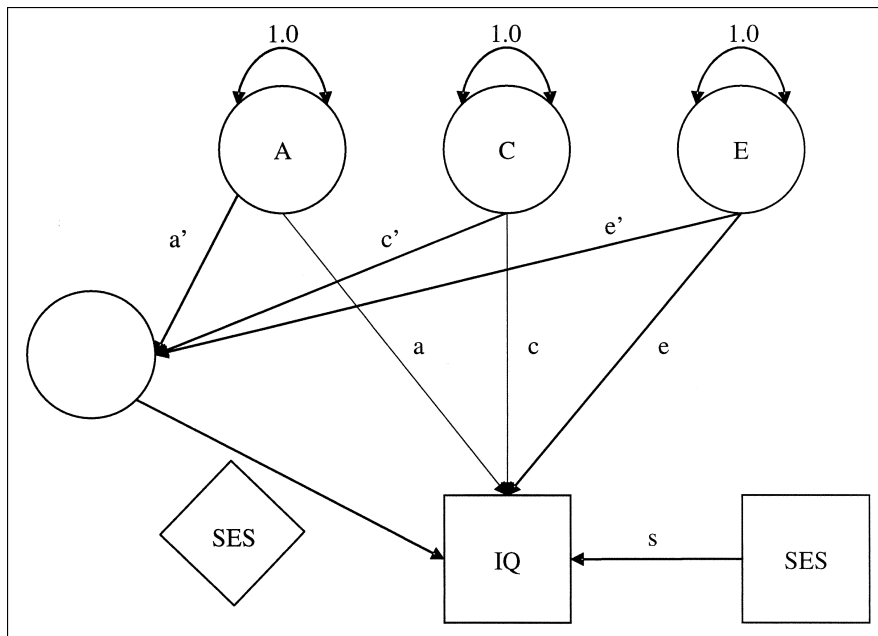

Fig. 1. Path diagram of the biometric model including interaction of biometric parameters with observed socioeconomic status (SES). A, C, and E are variances of additive genetic, shared environmental, and nonshared environmental components, respectively, and have been fixed at 1.0; a, c, and e are the main effects of A, C, and E on IQ; and a′, c′, and e′ are the interaction terms. The main effect of SES on IQ is denoted by s. The diamond containing SES and the empty circle are the Mx convention for denoting a latent variable interaction.

linear function of A, C, and E, we allowed each of the three components to interact linearly with observed SES, leading to the structural equation

$$IQ = sSES + (a + a'SES)A + (c + c'SES)C + (e + e'SES)E, \quad (1)$$

in which a, c, and e are the main effects of A, C, and E, respectively, and a', c', and e' represent the interaction of A, C, and E with SES. The s coefficient represents the main effect of SES on IQ. The model is illustrated in Figure 1. The model was fit using the Mx maximum likelihood modeling program (Neale, Boker, Xie, & Maes, 1999). A recent addition to Mx allows the user to place the value of an observed variable on a path, as indicated by the diamond denoting SES in Figure 1. Application of standard path-tracing rules to the path model in Figure 1 results in the structural equation specified by Equation 1. Mx uses maximum likelihood estimation to fit the parameters to the raw-data matrix (i.e., the observed IQ and SES scores of the twins, as opposed to their covariance matrix). Fitting to the raw-data matrix is a requirement for the interaction models we report, and results in degrees of freedom equal to the total number of observations (the number of participants times the number of variables per participant) minus the number of estimated parameters (Neale et al., 1999). Fit indices are therefore on a different scale from those based on traditional fitting to covariance matrices.

RESULTS

Results of the model fitting are given in Table 1. The parameters of the main-effects model and the interaction model are provided for the three tests (A, C, E, and main effect of SES for the main-effects model, and these parameters plus A', C', and E' interaction parameters for the interaction model).

The role of sampling error in the results can be assessed in three ways. With assumptions of normally distributed errors, the difference between the fits of the two models can be interpreted as a chi-square with three degrees of freedom arising from the addition of the three interaction parameters (see Table 2). The probabilities associated with these chi-square values assess the simultaneous contribution of the three interactions to the model. They suggest that the interactions jointly contribute a significant improvement in fit for FSIQ and PIQ, but not for VIQ.

A second method of assessing sampling error in our estimates is with confidence intervals on the individual main-effect and interaction parameters in the interaction model. These confidence intervals are also reported in Table 1. Mx computes them empirically by refitting the model, moving estimates from their optimized values until the chi-square fit function is increased by the appropriate amount (e.g., a chi-square change of 3.84 for a 95% confidence interval on a single parameter). This method has significant advantages over computing standard errors based on asymptotic distributional assumptions (Neale & Miller, 1997). The confidence intervals around most of the individual A', C', and E' interaction parameters for FSIQ and PIQ exclude zero. The tests of individual parameters are not as robustly significant as the simultaneous test of all three of them, however, suggesting that models in which one of the A', C', and E' interaction parameters equaled zero would fit nearly as well as a model with all three interactions.

Finally, sampling error can be assessed by the 95% confidence intervals around the results when the interaction is interpreted by computing the A, C, and E variance components for different levels of SES. We present these results next.

Table 1. *Results of model fitting, showing path coefficients (rows labeled A, C, E, A', C', E')*

Factor	Main-effects-model parameter	Interaction-model parameter	95% CI Lower	95% CI Upper
		FSIQ		
A	10.41	2.33	−7.28	9.67
C	11.66	16.83	11.43	21.87
E	9.14	13.32	10.92	15.68
SES	0.36	0.36	0.28	0.44
A'		0.16	0.05	0.29
C'		−0.12	−0.26	−0.01
E'		−0.1	−0.13	−0.05
		VIQ		
A	7.16	2.69	−5.06	7.98
C	8.56	12.03	7.48	16.32
E	6.97	7.42	5.66	9.32
SES	0.26	0.26	0.2	0.31
A'		0.1	−0.01	0.19
C'		−0.09	−0.22	0.02
E'		−0.01	−0.04	0.03
		PIQ		
A	−5.23	−2.26	−11.37	6.72
C	9.41	11.97	7.69	15.89
E	9.48	12.11	10.21	14.04
SES	0.22	0.22	0.16	0.28
A'		0.16	0.01	0.27
C'		−0.07	−0.19	0.03
E'		−0.06	−0.1	−0.02

Note. Additive variation in genotype, shared environment, and nonshared environment is denoted by A, C, and E, respectively. Their interactions with socioeconomic status (SES) are denoted by A', C', and E'. FSIQ = Full-Scale IQ; VIQ = Verbal IQ; PIQ = Performance IQ; CI = confidence interval.

The parameters in the main-effects model are interpreted as in any other twin-based model: The sum of their squares is equal to the phenotypic variance of the IQ score. The main-effect parameters in the interaction model are the estimated values of A, C, and E when SES equals zero, that is, in the most impoverished families. The interaction

Table 2. *Chi-square fits of the main-effects and interaction models and tests of the difference between the models*

Test	Fit Main-effects model	Fit Interaction model	Comparison χ^2 difference	p
Full-Scale IQ	8,236.4	8,220.6	15.8	.001
Verbal IQ	7,841.2	7,837.5	3.7	.3
Performance IQ	8,014.5	8,005.8	8.7	.034

parameters are difficult to interpret numerically because of the complex scaling issues introduced by the interactions with observed SES, measured on a scale from 0 to 100. The interactions are easiest to understand graphically, by plotting the variance in IQ accounted for by A, C, and E for different observed values of SES. The variance of each component (A, C, and E) of the phenotypic IQ variance is equal to the square of that component in Equation 1. For example, the IQ variance accounted for by additive genetic effects is given by

$$V_A = (a + a'SES)^2 \qquad (2)$$

Figure 2 presents the results of this analysis for FSIQ, VIQ, and PIQ. The figure shows 95% confidence intervals computed on the three variance components at intervals of SES ranging from 0 to 100. Note that the confidence intervals exclude a solution in which the individual variance components are constant across SES for A and E on FSIQ and for E on PIQ.

The percentage of the total IQ variance accounted for by each component (A, C, and E) is the ratio of the variance accounted for by that component to the total variance. For example, the percentage of IQ variability accounted for by genetic effects (the broad heritability) is given by

$$h_B^2 = \frac{(a + a'SES)^2}{(a + a'SES)^2 + (c + c'SES)^2 + (e + e'SES)^2} \qquad (3)$$

Figure 3 shows the FSIQ variance accounted for by the three components, with 95% confidence intervals. In the most impoverished families, the modeled heritability of FSIQ is essentially 0, and C accounts for almost 60% of the variability; in the most affluent families, virtually all of the modeled variability in IQ is attributable to A.

It is also informative to analyze the data more traditionally. We dichotomized the pairs into those above the median SES and those below the median SES and used random-effects analysis of variance (Guo & Wang, 2002) to estimate the intraclass correlations for MZ and DZ twins in the two SES groups. In the low-SES group, the intraclass correlation was .63 for DZ twins and .68 for MZ twins, consistent with h^2 of .10 and c^2 of .58; for the high-SES group, the DZ twin
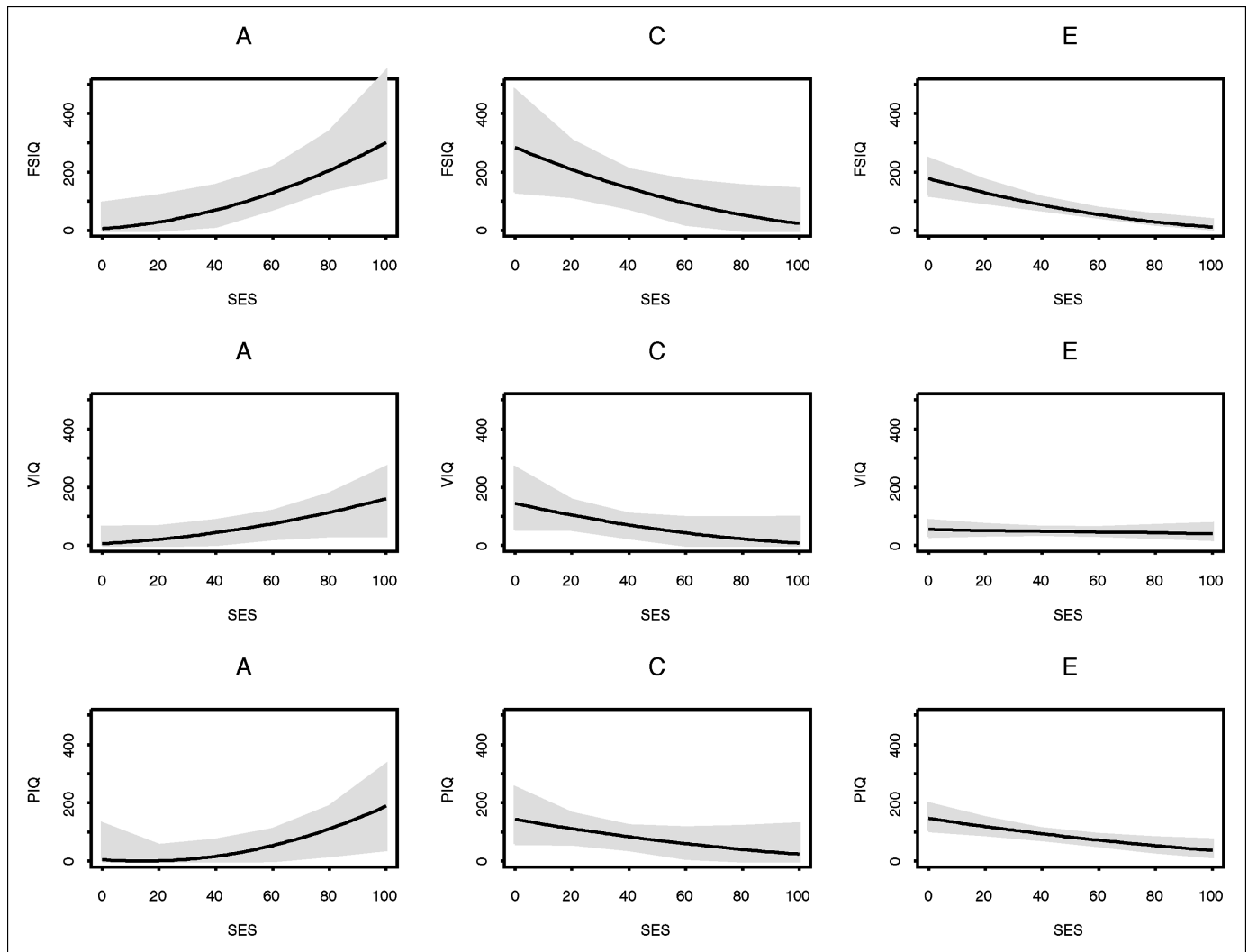

Fig. 2. Magnitude of the variance components A, C, and E for Full-Scale IQ (FSIQ), Verbal IQ (VIQ), and Performance IQ (PIQ) plotted as a function of observed socioeconomic status (SES). Shading indicates 95% confidence intervals.
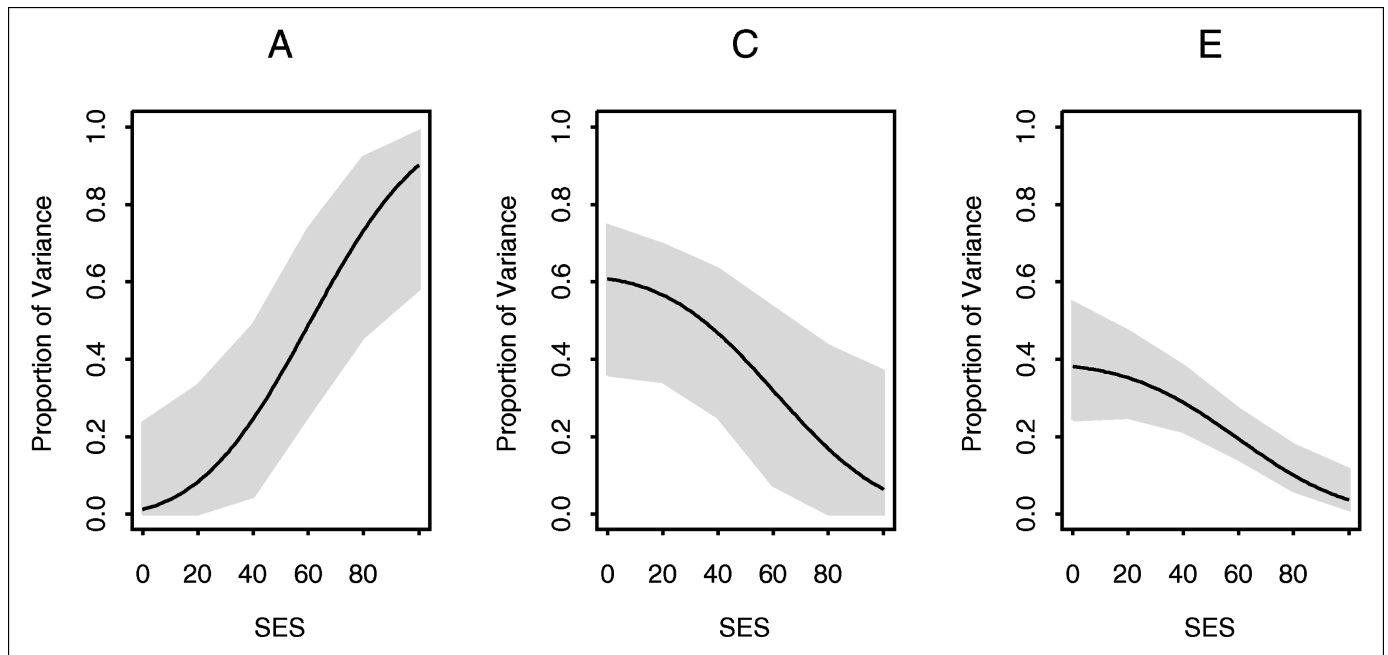
Fig. 3. Proportion of total Full-Scale IQ variance accounted for by A, C, and E plotted as a function of observed socioeconomic status (SES). Shading indicates 95% confidence intervals.

correlation was .51 and the MZ twin correlation was .87, consistent with h^2 of .72 and c^2 of .15.

DISCUSSION

These findings suggest that a model in which variability in intelligence among children is partitioned into independent components attributable to genes and environments is too simple for the dynamic interaction of genes and real-world environments during development. The relative importance of environmental differences in causing differences in observed intelligence appears to vary with the SES of the homes in which children were raised. SES is a complex variable, however, and the substantive interpretation to be placed on our results depends on an interpretation of what SES actually measures.

The most obvious interpretation of SES in this study is that it measured the quality of the environment in which the children were born and raised. Indeed, this is the function for which SES was intended. Under this interpretation, the observed interaction between SES and the biometric components of IQ could be indicative of precisely the kind of nonlinear relationship between rearing environment and intelligence that has been suggested by Scarr (1981) and Jensen (1981), with differences among poor environments contributing more to differences in phenotypic outcome than differences among middle-class or better environments contribute.

It would be naive, however, to interpret SES strictly as an environmental variable. Most variables traditionally thought of as markers of environmental quality also reflect genetic variability (Plomin & Bergeman, 1991). Children reared in low-SES households, therefore, may differ from more affluent children both environmentally and genetically (Gottesman, 1968), and the models we employed in this study do not allow us to determine which aspect of SES is responsible for the interactions we observed. Indeed, it will be difficult to separate

the genetic and environmental aspects of SES or other measures of the family environment in research designs of this kind, because children raised in the same home necessarily have the same SES.

Genetic variability in SES might also introduce a complication to the models themselves. Phenotypic SES and IQ are correlated, and that correlation is potentially mediated both genetically and environmentally. Therefore, the models are attempting to detect an interaction between genotype and environment in the presence of a correlation between genotype and environment, raising the concern that the presence of the correlation might introduce bias into the estimation of the interaction. However, Purcell (2003) has conducted an exhaustive series of simulations that suggest no bias is introduced, as long as the main effect of the moderating variable is included in the model, as we have done here. The presence in the model of the main effect of SES means that the biometric model fitting is actually being conducted on the portion of IQ that is independent of both the genetic and environmental components of SES. (We note, however, that omitting the main effect from the model did not change the results to a significant degree.)

The developmental mechanisms underlying the effect remain unclear. Although the models indicate that the A′, C′, and E′ interactions jointly contributed significant variance to differences in FSIQ and PIQ, the models were less able to distinguish which of the individual interactions with A, C, and E was most important in the effect. The interaction could be mediated primarily along genetic pathways, meaning that genetic differences among individuals are accentuated in favorable environments, as has been theorized by Bronfenbrenner (Bronfenbrenner & Ceci, 1994). It could also be that the slope of the IQ × Environment function is steeper at low levels of environment, as suggested by Scarr (1981) and Jensen (1981). Or maybe outcome simply becomes less predictable in poor environments, leading to an increase in E variability, as we have suggested (Turkheimer & Waldron, 2000) based on other evidence. To resolve this issue, it will be most important to study large

samples, which will provide greater power to discriminate differences among the genetic and environmental interactions.

It should also be noted that we found the interaction for FSIQ and PIQ only; the interaction was in the predicted direction for VIQ but did not approach statistical significance. Specifying more precisely the kinds of abilities for which the interaction is likely to occur is the current focus of investigation in our laboratory. In the National Collaborative Perinatal Project, scores are available for the individual subtests of the WISC, and for a variety of other ability and achievement measures. We are developing models of interactions between SES and genetic and environmental variance in multivariate analyses of factor scores based on correlations among individual tests.

In the fractious history of scientific investigations of the heritability of intelligence, the effects of poverty, and the relations between them, there has been only one contention with which everyone could agree: Additive models of linear and independent contributions of genes and environment to variation in intelligence cannot do justice to the complexity of the development of intelligence in children. Only recently have statistical models and computational capacity advanced to the point that less simplistic models can actually be fit. Although there is much that remains to be understood, our study and the ones that have preceded it have begun to converge on the hypothesis that the developmental forces at work in poor environments are qualitatively different from those at work in adequate ones. Clarification of the nature of these differences promises to be a fascinating, and hopefully unifying, subject for future investigation.

REFERENCES

Bronfenbrenner, U., & Ceci, S.J. (1994). Nature-nurture reconceptualized in developmental perspective: A bioecological model. *Psychological Review, 101*, 568–586.

DeFries, J.C., & Fulker, D.W. (1985). Multiple regression analysis of twin data. *Behavior Genetics, 15*, 467–473.

Eaves, L.J., & Jinks, J.L. (1972). Insignificance of evidence for differences in heritability of IQ between races and social classes. *Nature, 240*, 84–88.

Fischbein, S. (1980). IQ and social class. *Intelligence, 4*, 51–63.

Gottesman, I.I. (1968). Biogenetics of race and class. In M. Deutsch, I. Katz, & A. Jensen (Eds.), *Social class, race, and psychological development* (pp. 11–51). New York: Holt, Rinehart & Winston.

Guo, G., & Wang, J. (2002). The mixed or multilevel model for behavior genetic analysis. *Behavior Genetics, 32*, 37–49.

Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: Free Press.

Jensen, A.R. (1981). Raising the IQ: The Ramey and Haskins study. *Intelligence, 5*, 21–40.

LaBuda, M.C., & DeFries, J.C. (1990). Genetic etiology of reading disability: Evidence from a twin study. In G.V. Pavlidis (Ed.), *Perspectives on dyslexia* (Vol. 1, pp. 47–76). New York: Wiley.

McGue, M., Bouchard, T.J., Iacono, W.G., & Lykken, D.T. (1993). Behavioral genetics of cognitive ability: A life-span perspective. In R. Plomin & G.E. McClearn (Eds.), *Nature, nurture & psychology* (pp. 59–76). Washington, DC: American Psychological Association.

Myrianthopoulos, N.C., & French, K.S. (1968). An application of the U.S. Bureau of Census socioeconomic index to a large diversified population. *Social Science and Medicine, 2*, 283–299.

Neale, M., Boker, S., Xie, G., & Maes, H. (1999). *Mx: Statistical modeling.* (Available from Medical College of Virginia, Department of Psychiatry, Box 126 MCV, Richmond, VA, 23298)

Neale, M.C., & Miller, M.B. (1997). The use of likelihood-based confidence intervals in genetic models. *Behavior Genetics, 27*, 113–120.

Nichols, P., & Chen, T. (1981). *Minimal brain dysfunction: A prospective study.* Hillsdale, NJ: Erlbaum.

Plomin, R. (1999). Genetics and general cognitive ability. *Nature, 402*, C25–C29.

Plomin, R., & Bergeman, C.S. (1991). The nature of nurture: Genetic influence on "environmental" measures. *Behavioral and Brain Sciences, 14*, 373–427.

Plomin, R., DeFries, J.C., McClearn, G.E., & McGuffin, P. (2000). *Behavioral genetics.* New York: W.H. Freeman.

Purcell, S. (2003). Variance components models for gene–environment interaction in twin analysis. *Twin Research, 5*, 554–571.

Rowe, D.C., Jacobson, K.C., & van den Oord, E. (1999). Genetic and environmental influences on vocabulary IQ: Parental education level as moderator. *Child Development, 70*, 1151–1162.

Scarr, S. (1981). *Race, social class, and individual differences in I.Q.: New studies of old issues.* Hillsdale, NJ: Erlbaum.

Scarr-Salapatek, S. (1971). Race, social class, and IQ. *Science, 4*, 1285–1295.

Stoolmiller, M. (1999). Implications of the restricted range of family environments for estimates of heritability and nonshared environment in behavior-genetic adoption studies. *Psychological Bulletin, 125*, 392–409.

Thompson, L.A., Tiu, R.D., & Detterman, D.K. (1999, July). *Differences in heritability across levels of father's occupation.* Poster presented at the annual meeting of the Behavior Genetics Association, Vancouver, British Columbia, Canada.

Turkheimer, E. (1991). Individual and group differences in adoption studies of IQ. *Psychological Bulletin, 110*, 392–405.

Turkheimer, E., & Waldron, M.C. (2000). Nonshared environment: A theoretical, methodological and quantitative review. *Psychological Bulletin, 126*, 78–108.

U.S. Census Bureau. (2002). *Historical poverty tables.* Retrieved July 31, 2003, from http://www.census.gov/hhes/poverty/histpov/hstpov1.html

van den Oord, E., & Rowe, D.C. (1997). An examination of genotype-environment interactions for academic achievement in an U.S. longitudinal survey. *Intelligence, 25*, 205–219.