Supplementary Appendix 1

Below is a fuller outline of the analysis techniques used in the main text, including the regression equations, bootstrap methods and correction of s.e. for double entry of the data.

**Zygosity Determination in the GHCA samples:** Western Reserve Reading Project – DNA analysis by buccal swab procedure. Twins Early Development Study - parent questionnaire of physical similarity, with follow-up DNA testing in cases where zygosity was unclear. Minnesota Center for Twin and Family Research - consensus of four indicators: a standard zygosity parent questionnaire; staff perception of physical similarity; an algorithm based on ponderal index, cephalic index and fingerprint ridge count. DNA analysis of 12 blood group anitigens from blood samples was performed if there was any discrepancy among these three methods. Zygosity was determined in almost all cases using a panel of DNA markers. Twin Cognition Study – Cross-checked DNA markers, blood group results and phenotypic data. Netherlands Twin Register – typing of DNA or blood group polymorphisms.

**Twin Methodology and DeFries-Fulker Regression:** DeFries-Fulker regression analysis (LaBuda, DeFries and Fulker, 1986) uses monozygotic (MZ; genetically identical) and dizygotic (DZ or fraternal; sharing 50% of genetic variation on average) twin pairs, regressing twin two's score (C) on twin one's scores (P) and the coefficient of relationship (R; 1 for MZ and .5 for DZ pairs). A third term estimating the interaction between the P and R, yields direct estimates the heritability (proportion of sample variance accounted for by genetic influences; $h^2$) and the proportion of variance accounted for by family-wide environmental influences ($c^2$). All twin pairs were double-entered, with the twin assignment reversed in the second entry (i.e. twin 1 becomes twin 2 and *vice versa*). In equation 1, $\beta_1$ estimates $c^2$ and $\beta_3$ estimates $h^2$ when the data are suitably transformed. *K* is a constant:

$$C = \beta_1 P + \beta_2 R + \beta_3 PR + K \quad (1)$$

Further extensions to this equation can test changes in the estimations of $h^2$ and $c^2$ according to IQ score:

$$C = \beta_1 P + \beta_2 R + \beta_3 PR + \beta_4 P^2 + \beta_5 P^2 R + K \quad (2)$$

In equation 2, $\beta_4$ measures the linear relationship between twin 1's IQ score and the predictability of twin 2's IQ score from twin 1's, independent of genetic relationship. This tests for the linear change in $c^2$ as twin 1's score increases. $\beta_5$ measures how this variable differs as a function of the relationship between twin 1 and twin 2 and is the corresponding test for linear change in $h^2$ (Cherny, Cardon, Fulker &DeFries, 1992). In the cross-sectional GHCA sample, we first applied equation 1 and 2 to the full sample collectively then extended the analysis to examine the effect of age on the estimates derived. The sample was split into 3 age groups, childhood ( aged 4-12; n pairs = 6044), adolescence (aged 12-18; n pairs = 4304) and adulthood ( ages 18+; n pairs = 549) and orthogonal linear and quadratic contrast codes (*age_lin* and *age_quad*) were constructed and allowed to interact with all the terms in equation 2. In such a regression, $\beta_9 P^2$*age_lin* and $\beta_{10} P^2 R$*age_lin* test for a difference between the child and adult age groups on the relationship between ability and the estimate of $c^2$ and $h^2$ respectively. $\beta_{16}$ $P^2$*age_quad* and $\beta_{17} P^2 R$*age_quad* compare the estimates for these groups collectively to those for the adolescent group . Applying equation 2 separately for each age group provided estimates for the

ability-dependent terms at each age. Additionally, splitting the sample into 6 groups by median splitting each age group on IQ score and applying equation 1 separately for each of these subsamples gave estimates of $c^2$ and $h^2$ for the top and bottom halves of the ability distribution in each age goup separately. For this analysis pairs were only double entered if both twins met criteria for the ability cut-off. All reported *p* values for the GHCA sample are derived by bootstrapping the regression estimates in the following way: Twin pairs were sampled at random, with replacement, from a single entered dataset and twin assignment was randomized for each pair. All regressions described were performed on the resulting data and the coefficents saved. This process was repeated 10,000 times. The standard deviations of the resulting betas were used as s.e.s of the estimates and *p* values we derived by centring the distribution around zero and calculating twice the proportion of the estimates that exceeded the observed value.

In the longitudinal sample, the coefficient of relationship (R) took on the value of 1.0 for MZ and .5 for DZ twin pairs as before. For biological siblings the value of .5 for also used, as they are genetically as similar as DZ pairs. Adoptive sibling pairs took a value of 0.0, as they are not genetically related. An extra variable, the age gap between siblings in days (0 for all twin pairs), was added to equations 1 and 2 and allowed to interact with the estimates of $c^2$ and $h^2$: All reported parameters are derived from regressions with age gap included as a moderator. This better enabled comparison of estimates between this sample and the cross-sectional twin study presented above (the estimates for when age gap = 0 to be to most accurate estimate of the maximum effect of shared environmental factors, the influence of which will diminish as age gap increases). If this difference is not modeled the smaller age gap between all biologically related pairs compared to the adoptive siblings will overestimate heritability as the increased correlation, which can reasonably be attributed to both increased environmental sharing in the biological siblings and to genetic influence, will be attributed in the regression purely to genetic influence. Standard errors of the regression estimates are calculated using the method outlined by Koehler and Rodgers (Kohler & Rodgers, 2001) by altering the `robust cluster` command in STATA (StataCorp., 2007) to give robust standard errors using *n* instead of *n - k -1* as a multiplier (for details see http://www.ssc.upenn.edu/~hpkohler/data-and-programs/twdfeff/twdfeffprograms.html#x1-130005). This method of calculation accounts for the fact that the double entered data are only independent at the level of the twin pair.

References


1. LaBuda, M.C., DeFries, J.C. & Fulker, D.C., (1986). Multiple regression analysis of twin data obtained from selected samples. *Genetic Epidemiology,* **3**, 425-433.

2. Cherny, S.S., Cardon, L.R., Fulker, D.W. & DeFries, J.C. (1992). Differential heritability across levels of cognitive ability. *Behavior Genetics,* **22**, 153-162.

3. Kohler, H.P. & Rodgers, J.L. (2001). DF-analyses of heritability with double-entry twin data: Asymptotic standard errors and efficient estimation, *Behavior Genetics* **31**, 179–191.


Supplementary Appendix 2

These supplementary materials outline the tests of validity performed on the LTS study sample in an attempt to ensure that the apparent prolonged critical period found in both our samples could not be readily explained by any confounding variables.

**Assortative mating in parents of the LTS twins:** We tested for patterns of assortative mating for IQ in the parents of the LTS twins. Assortative mating increases between-family variability and therefore manifests as $c^2$ in the twin design. If assortative mating was higher among higher IQ parents, then this would be a potential explanation for the increased $c^2$ seen in higher IQ individuals at age 16.

Full-scale WISC-R IQ scores were assessed in parents at the time of intake of the family into either the Twin Infant Project (TIP) or the Longitudinal Twin Study (between 3 and 14 months post partum; the two studies were amalgamated to construct the current LTS). In the event that this information was given twice during this period, the responses were averaged within parent. If information was available from just one parent, this value was used alone. In total, data was available for 400 fathers and 447 mothers, for a total of 399 families with complete parental information. The mean IQ score for mothers was 104.91 (sd = 12.37) and for fathers 107.50 (sd = 12.89). The correlation between parental scores was $r(399) = .388$, $p < .001$. This demonstrates a moderate amount of assortative mating for IQ in the parents of the LTS twins, which could account for some of the variance attributed to c2. In order to assess whether the extent of assortative mating was different depending on the ability level of the parents, we correlated mean parental IQ and absolute difference between parental scores. The average difference was 11.26 points (sd = 8.68). The correlation between mean parental score and the this difference score was $r(399) = .155$, $p = .002$. This suggests that assortative mating was significantly *less* strong as mean parental IQ score increased. For this reason, patterns of assortative mating cannot account for our results.

**Measurement variance in the IQ scale:** The second test we conducted was to examine the factor structure of the first unrotated principal component of the intercorrelations among the subtests of the WAIS-III assessed at age 16 in the LTS sample. If the factor structure differs according the ability level then it follows that the measure of full-scale IQ is actually measuring something different depending on the ability of the individual being assessed. This could potentially be a confounding factor in assessing differences in etiological influences depending on IQ score. This has been demonstrated inconsistently in tests of Spearman's Law of Diminishing Returns(Spearman, 1927), which posits that "*The correlations [between different tests] always become smaller—showing the influence of g on any ability to grow less—in just those classes of person which, on the whole, possess this g more abundantly. The rule is, then, that the more 'energy' [i.e., g] a person has available already, the less advantage accrues to his ability from further increments of it*" (p. 219). To test this hypothesis we followed the methods outlined by Jensen (2003). First we extracted the first unrotated principal component from the 11 subtests for the entre LTS sample measured at age 16 and then successively for those in the sample with IQs measured to be above and below 100 (the population average). Table S1 demonstrates the factor loadings derived from this analysis as well as the proportion of variance explained.

TABLE S1: *Factor loadings for the 11 subtests of the WISC-III for the full sample and for the two ability subsamples*

| Age 16 WAIS-III sub-test | loading on first principal component | Loading full-scale IQ >100 | loading full-scale IQ < 100 |
|---|---|---|---|
| Vocabulary | .84 | .78 | .81 |
| Similarities | .71 | .58 | .65 |
| Arithmetic | .71 | .53 | .52 |
| Digit Span | .38 | .09 | .22 |
| Information | .82 | .77 | .72 |
| Comprehension | .78 | .68 | .74 |
| Picture Completion | .39 | .06 | .00 |
| Digit Symbol | .42 | .20 | .17 |
| Block Design | .65 | .34 | .19 |
| Picture Arrangement | .35 | -.06 | .07 |
| Object Assembly | .52 | .20 | .13 |
| **Variance explained by first principal component:** | 38.80% | 22.68% | 23.37% |

It can be seen that, although all the factor loadings across groups follow the same pattern, the loading and the variance explained are lower overall for the two truncated groups. This is due to the restricted range in the scores resulting from selecting on full-scale IQ score. To compare the factor structure between the two ability groups, we calculated the average intercorrelation for each group using Kaiser's (1968) formula in which the eigenvalue of the first principal component -1 is divided by the number of variables -1 The values were .14 and .16 for above a below 100 respectively. We then used Fisher's (1915) r-to-z transformation to test for a significant difference between the two values. For this contrast $z' = -.20$, $p = .83$. There is therefore no evidence that the proportion of variance captured by the principal component differs between ability levels. Additionally, the congruence coefficient between the factors for the two subsamples was found to be .99, demonstrating the extreme similarity between the factor loadings.

**Gene x Environment Interactions:** Finally we wanted to rule out unmeasured gene x measured environment interactions as an explanation for our results. Previous studies have shown that the heritability of IQ is moderated by both years of parental education and socioeconomic status (Rowe, Jacobson, Van den Oord, 1999; Turkheimer, Haley, Waldron, D'Onofrio& Gottesman, 2003). In the LTS sample we were able to test the moderating effect of both parental education and parental IQ score on the etiology of age sixteen IQ (both are predictors of socio-economic status). The method of analysis used was the moderated paths variance components model outlined in Purcell (2002) (Figure S1). In traditional biometric models, variance is partitioned into proportion explained by additive genetic ($a^2$), shared environmental ($c^2$) and unique environmental influences ($e^2$), modeling the expected covariance between twins. The moderated paths model adds a continuous moderation of these proportions by an environmental variable. The mean of the trait of also moderated by the environmental variable, which removes the shared variance between the moderator and the trait from the covariance model, meaning that any detected interaction will be between the moderator and variance specific to the trait. This removes any confounding effect of gene-environment correlations (shared genetic influences between

the trait and the moderator). In the resulting model the expected mean of the trait T in twin *i* is $\mu + \beta MM_i$ and the expected trait variance is $Var(T_i) = (a + \beta_x M_i)^2 + (c + \beta_Y M_i)^2 + (e + \beta_Z M_i)^2$.

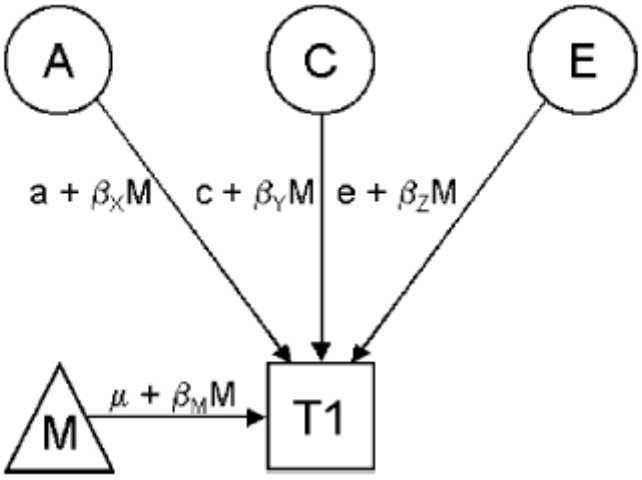


FIGURE S1: *Path diagram for one twin in the GxE interaction model. a, c, e = unmoderated additive genetic, shared environmental and unique environmental influences.* $\beta_x$; $\beta_Y$ and $\beta_Z$ *= moderated components of a, c and e, respectively.* $\beta_M$ *= main effect of moderator; M = moderator;* $\mu$ *= grand mean.*

Seven variables are therefore estimated in the model: the unmoderated components a,c and e, moderated components $\beta_x$, $\beta_Y$ and $\beta_Z$ and main effect $\beta_M$. Parameters can be dropped successively from the model and -2 log likelihoods (-2LL) can be compared to that of the full model to determine the best-fitting model (the difference between the -2LL has a $\chi^2$ distribution with degrees of freedom being $\Delta$df between the two models).

Years of education was self-reported by parents at the time of entry of the families into the study and information was available for both parents in the majority of families (n = 452). For some families, data for reported more than once at different times and in these cases the mean of the two scores was used. To construct the variable used in our analyses maternal and paternal years of education was calculated, or if information was only available from one parent this one data point was used (n = 12). Resulting scores were standardized. Mean maternal years of education was 14.26 years (sd = 2.29) and paternal was 14.58 years (sd = 2.52). Full-scale WISC-R scores were used as the moderating IQ variables (details above) which were also standardized. The correlation between twin one's age 16 IQ and parental education was .40 and between twin one's age 16 IQ and parental IQ was .51. Twin correlations above and below the median on parental IQ and parental years of education for both raw age 16 IQ scores and residuals from regressing the parental variables predictors on age 16 IQ are presented in Table S2.

TABLE S2: *MZ and DZ twin intra-class correlations for age 16 IQ in the LTS as a function of parental environmental variables, with and without residualization for shared variance between the two variables*

| | parental education | | parental IQ | | parental education (residualized) | | parental IQ (residualized) | |
|---|---|---|---|---|---|---|---|---|
| **cut-off** | MZ | DZ | MZ | DZ | MZ | DZ | MZ | DZ |
| above median | 0.82 | 0.43 | 0.78 | 0.41 | 0.8 | 0.39 | 0.74 | 0.4 |
| below median | 0.8 | 0.52 | 0.78 | 0.41 | 0.78 | 0.48 | 0.76 | 0.2 |
| total sample | 0.84 | 0.51 | 0.83 | 0.5 | 0.79 | 0.43 | 0.75 | 0.34 |

Table S3 reports model fit statistics using parental education as a moderator of age 16 IQ etiology. It can be seen that the moderation of the estimates of a2, c2 and e2 by a parental education can be dropped from the model with no decrement in fit. The mean moderation, however, cannot be dropped. The final model gives estimates of $a^2 = .72$, $c^2 = .07$ and $e^2 = .21$.

TABLE S3: *Model fit statistics for the moderation model of parental education on age 16 IQ etiology*

| model | -2LL | df | Δχ² | Δdf | p | AIC |
|---|---|---|---|---|---|---|
| full ACE-XYZ-M | 1831.7 | 392 | | | | |
| drop XYZ | 1832.33 | | 0.63 | 3 | 0.89 | -5.373 |
| drop M | 1911.97 | | 80.97 | 1 | <.001 | 78.973 |

Table S4 reports model fit statistics using parental IQ as a moderator of age 16 IQ. We found that moderation of c2 and e2 by parental IQ score could be dropped from the model without a decrement in fit, but moderation of the mean of age 16 IQ and moderation of a2 could not. The estimates for etiological influences as a function of standardized parental IQ score are displayed in figure S2.

Table S4: *Model fit statistics for the moderation model of parental IQ on age 16 IQ etiology*

| model | -2LL | df | Δχ² | Δdf | p | AIC |
|---|---|---|---|---|---|---|
| full ACE-XYZ-M | 1693.21 | 378 | | | | |
| drop XYZ | 1706 | | 12.79 | 3 | <.001 | 6.793 |
| drop Y and Z | 1693.61 | | 0.4 | 2 | 0.818 | -3.6 |
| drop M | 1825.06 | | 131.85 | 1 | <.001 | 129.8 |

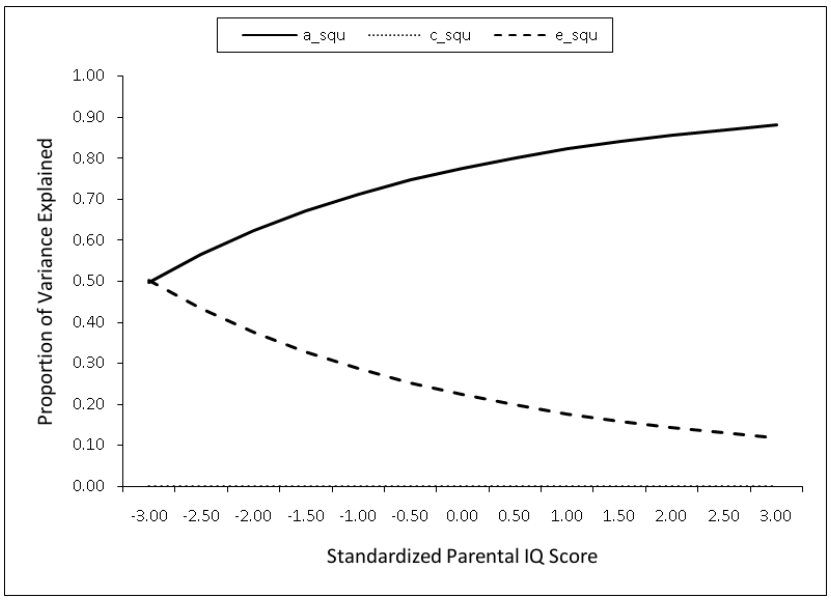


Figure S2: *Etiological influences on age 16 IQ as a function of parental IQ score*

It can be seen that the magnitude of additive genetic influences increases as parental IQ increases, with the relative influence of the unique family environment decreasing. When the shared variance between parental IQ and age 16 IQ is controlled for, there is no influence of the shared family environment at any level of parental education. Importantly, the influence of parental IQ on heritability is in the *opposite* direction to that of the cotwin's IQ score, in which heritability decreases as score increases, and so cannot account for that result.

References


1. Spearman, C.E., (1927). *The abilities of man.* Macmillan, London.
2. Jensen, A.R., (2003). Regularities in Spearman's law of diminishing returns, *Intelligence* **31**, 95-105.
3. Kaiser, H. (1968). A measure of the average intercorrelation. *Educational and Psychological Measurement,* 28:245-247.
4. Fisher, R.A. (1915). Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. *Biometrika*, **10**, 507–521.
5. Rowe, D.C., Jacobson, K.C., Van den Oord, E.J.C.G., (1999). Genetic and environmental influences on vocabulary IQ: Parental education level as moderator. *Child Development,* **70**, 1151-1162.
6. Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B. & Gottesman, I.I., (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science,* **14**, 623-628.

7. Purcell, S. (2002). Variance components model for gene-environment interaction in twin analysis. *Twin Research and Human Genetics,* **5**, 554-571.