Supplementary Online Material to **“**Replication of a Gene-Environment Interaction via Multimodel Inference: Additive-Genetic Variance in Adolescents’ General Cognitive Ability Increases with Family-of-Origin Socioeconomic Status” by R. M. Kirkpatrick, M. McGue, & W. G. Iacono

Supplementary Methods

Sample

MTFS is a longitudinal study of a community-based sample of same-sex twins, born between 1972 and 1994 in the State of Minnesota, and their parents. SIBS is an adoption study of sibling pairs and their parents; its community-based sample includes families in which both siblings are adopted, in which both are biologically related to the parents, or in which one is adopted and one is biologically related. Per the SIBS inclusion criteria, any sibling in the sample who was adopted into the family will not be biologically related to his or her co-sibling. For adopted siblings, the mean age at placement was 4.7 months (SD = 3.4 months).

Between MTFS and SIBS, 2,504 families have visited the Minnesota Center for Twin & Family Research. For the purposes of our analyses, the primary sample (*N* = 4,973 individuals in 2,494 families) comprises five distinct family-types:

1. Monozygotic- (MZ) twin families (*n* = 2401, in 1204 families),
2. Dizygotic- (DZ) twin families (*n* = 1348, in 675 families),
3. SIBS families with two adopted offspring (*n* =567, in 285 families),
4. SIBS families with two biological offspring (*n* = 415, in 208 families),
5. “Mixed” SIBS families with 1 biological and 1 adopted offspring (*n* = 242, in 122 families).

Descriptive characteristics of the sample are presented in Table S1. Based on self-reported ethnicity, this sample is predominantly Caucasian of European ancestry (“White,” 84.4%). Most of the adoptees were adopted internationally from Korea, so 10% of the sample is Asian. The remaining 5.6% of the sample report their ancestry as other than White or Asian. In addition to this primary sample, one of our secondary analyses used a sample of 3,916 parents from MTFS and SIBS.

The MTFS sample is composed of two cohorts, an eleven-year-old cohort (10-13 years old at intake; mean age = 11.78) and a seventeen-year-old cohort (16-18 years old at intake; mean age = 17.47). The age range of the siblings at intake was 10-19 for the younger sibling of each pair, and 12-20 for the older.

Measurements

Zygosity.

Twin zygosity was assessed at intake using three indicators: a standard, parent-completed zygosity questionnaire, staff judgment of physical similarity, and an algorithm that used various anthropometric measures. Zygosity determination for putatively DZ twins who provided DNA samples has subsequently been verified from genome-wide marker data (Miller et al., 2012).

Cognitive ability.

Measurement of GCA was included in the design of the intake assessment for most participants, by way of an abbreviated form of the Wechsler Intelligence Scale for Children-Revised (WISC-R) or Wechsler Adult Intelligence Scale-Revised (WAIS-R), as age-appropriate (that is, 16 or younger, and older than 16, respectively). The short forms consisted of two Performance subtests (Block Design and Picture Arrangement) and Verbal subtests (Information and Vocabulary), the scaled scores on which were prorated to determine Full-Scale IQ (FSIQ).

FSIQ estimates from this short form have been shown to correlate 0.94 with FSIQ from the complete test (Sattler, 1974). Parents in the SIBS sample were an exception, in that they were not tested with this short form of WAIS-R until the first SIBS follow-up assessment. By design, only one parent per SIBS family returned for this follow-up, which was usually the mother. As a result, IQ data for SIBS fathers is very limited in its availability.

Supplementary Notes

Note #1: Early Studies of SES and GCA

Scarr (Scarr-Salapatek, 1971) was the first to investigate whether the heritability of children's GCA might vary as a function of their family SES, using a sample of ~1000 twin pairs from the Philadelphia public school system. Scarr's results were consistent with higher heritabilities at higher SES levels. But, some (e.g., Turkheimer et al., 2003; Hanscombe et al., 2012) have criticized this early study for its major limitations, such as its indirect resolution of zygosity in same-sex twin pairs, and the use of neighborhood census-tract data (rather than data from the twins' actual parents) to operationalize SES. A later study of Swedish twins (Fischbein, 1980) did not suffer from these limitations. Fischbein stratified twins by their father's occupational status (3 levels) and reported the MZ and DZ intraclass correlations by stratum. These results were also consistent with higher heritabilities at higher SES levels, though the sample size was rather small (<300 twin pairs), and Fischbein reported no inferential statistics. More importantly, results were only presented for two ability tests, separately, and not for any composite from multiple tests, so the twins should not be considered assessed for *general* cognitive ability.

Subsequently, Rowe and co-authors reported two studies (van den Oord & Rowe, 1998; Rowe, Jacobson, & van den Oord, 1999) investigating gene-environment interaction in cognitive abilities. Unlike those of Scarr and Fischbein, both studies employed data-analysis methods allowing statistical inference about moderation effects. van den Oord and Rowe utilized a genetically informative design by identifying pairs of siblings, half-siblings, and cousins in the National Longitudinal Survey of Youth. Their sample comprised ~3300 children who had been assessed with an achievement test composed of a mathematics, word-recognition, and readingcomprehension subtest. Participants' home environments had been measured on eleven variables, including family poverty status and each parent's educational attainment. Multilevel regression provided little evidence that the eleven variables moderated heritability of total achievement scores or subtest scores. Instead, most significant (at $p < 0.01$) interactions indicated that either total variance, or unshared-environmental variance only, were smaller in favorable home environments compared to less-favorable home environments. In contrast, Rowe et al. reported that parental education significantly moderated both heritability and shared-environmentality of picture-vocabulary scores in their sample of ~1900 secondary-school-age sibling pairs from the Add Health dataset. Their double-entered DeFries-Fulker regression showed that heritability increased, and shared-environmentality decreased, with increasing parental education. However, neither of these studies assessed their participants on a broad spectrum of abilities. The achievement test in van den Oord & Rowe's study, of course, only represents abilities heavily dependent on formal education, whereas Rowe et al.'s participants only took a Vocabulary test. These instruments should be regarded as measuring a narrower construct than GCA.

Note #2: Details Concerning Attempted Replications of Turkheimer et al. (2003)

We are aware of five studies of GCA interpretable as attempts at replicating Turkheimer et al.'s (2003) $A \times SES$ and $C \times SES$ effects (Harden, Turkheimer, & Loehlin, 2007; van der Sluis, Willemsen, de Geus, Boomsma, & Posthuma, 2008; Grant, Kremen, Jacobson, Franz, Xian, et al., 2010; Hanscombe et al., 2012; Bates, Lewis, & Weiss, 2013). Each of these studies applied the Purcell (2002) continuous-moderator model to scores on a test composed of both verbal and non-verbal subtests. Thus, we do not consider studies of SES-moderation in specific abilities, such as reading achievement (e.g., Kremen et al., 2005) or mathematics achievement (e.g., Tucker-Drob & Harden, 2012), or separate verbal and non-verbal ability factors (e.g., Asbury, Wachs, & Plomin, 2005). We also do not consider studies of cognition in very young children (e.g., Tucker-Drob, Rhemtulla, Harden, Turkheimer, & Fask, 2011), because cognitive tests prior to age 2 correlate weakly with cognitive ability measured at later ages (Wilson, 1983) and sample a limited amount of verbal content (Bouchard & McGue, 2003), rendering them questionable as measures of the GCA construct. General cognitive ability does not emerge as a stable trait until around age 4 or 5.

Harden et al. (2007) reported an analysis using a sample of 839 pairs of adolescent twins who sat for the 1962 National Merit Scholarship Qualifying Test (NMSQT). The NMSQT, an aptitude test designed to assess scholarship candidates' potential for success in future educational endeavors, comprised five subtests: English Usage, Mathematics Usage, Social Science Reading, Natural Science Reading, and Word Usage / Vocabulary. Evidently, the NMSQT sampled relatively little content disconnected from mastery of a formal educational curriculum or involving nonverbal cognitive processes. It is therefore with some reservation that we regard it as a measure of GCA. Indeed, Harden et al. acknowledged that the NMSQT is not an IQ test like that used in Turkheimer et al. (2003).

The actual phenotype biometrically decomposed in Harden et al.'s (2007) analysis was the latent common factor extracted from the five NMSQT subtests. Harden et al. used two SES variables, mid-parental education level and family income. A comparison of the observed income distribution to that of the 1960 U.S. census revealed that their twins disproportionately came from relatively affluent families. Harden et al. conducted separate moderation analyses with each SES variable, and in both analyses estimated $A \times SES$ and $C \times SES$ effects. The analyses yielded a significant ($p \approx 0.02$) effect, in the hypothesized direction, of income on additive-genetic influence only, but also suggestive evidence of the other moderation effects (all other ps < 0.1). However, none of the point estimates of the $C \times SES$ effect were in the hypothesized direction. Thus, Harden et al. (2007) constitutes a replication of the $A \times SES$ effect of Turkheimer et al. (2003): increasing additive-genetic variance with increasing SES.

In a follow-up to Harden et al. (2007), Loehlin et al. (2009) evaluated the consequences of assortative mating. Through simulations, they showed that income's estimated $A \times SES$ effect changed little in value by manipulating SES-independent assortative-mating parameters. Further, they showed that the real-data DZ-twin correlations, and spousal correlations for education level, did not vary by family income levels, which suggests that SES-dependent assortative mating is unlikely to be present in their sample.

Van der Sluis et al. (2008) reported a replication attempt in an adult sample of Dutch twins. The sample comprised two age cohorts: a younger cohort (mean age ~27, $N = 385$) and an older cohort (mean age ~49, $N = 370$). Participants were assessed on an abbreviated form of the Dutch WAIS-III consisting of 9 subtests—a broad sampling of the domain of cognitive tasks. FSIQ scores were age- and sex-corrected prior to analyses. Midparental education level was scored as a binary indicator for completing higher education; on average, twins in the younger cohort had more highly-educated parents. Because of this cohort difference (and because of sex differences with respect to other variables not germane to the present discussion), van der Sluis et al. fit moderation models separately by cohort and sex. All biometric-moderation effects of parental education could be dropped without significantly disturbing model fit, apart from that for shared-environmental variance among older males. This effect size translates into $c^2 = 0.47$ among older males with higher-educated parents, but $c^2 = 0$ otherwise. Clearly, the results of van der Sluis et al. (2008) did not replicate Turkheimer et al. (2003).

Grant et al. (2010) conducted a replication attempt with a sample of ~6000 adult-male twins from the Vietnam-Era Twin Registry, whose general cognitive ability had been assessed with the Armed Forces Qualification Test upon their induction to the military. Midparental education level was calculated for each pair as the average of the twins' reports of their mother's and father's years-of-education completed. Then, these averages were converted to a scale on which the lowest level of parental education was coded 0, and the highest level was coded 1. But, moderation analysis provided no evidence that any of the three biometric variance components varied as a function of parental education. Thus, Grant et al. (2010) also did not replicate the Turkheimer et al. (2003).

Hanscombe et al. (2012) attempted to replicate Turkheimer et al. (2003) in a longitudinal dataset from the Twins Early Development Study. The sample was composed of 8716 British twin pairs in which at least one twin had been IQ-tested at any of the target ages (2, 3, 4, 7, 9, 10, 12, and 14; assessments differed appropriately by age), and for which at least one of the three indices of parental SES was available. These three indices were (1) parental education level and occupational status, as recorded at recruitment, when the twins were 18 months old; (2) parental education level and occupational status as recorded when the twins were 7 years old; and (3) parental income as measured when the twins were 9 years old. Altogether, Hanscombe et al. presented results of separate biometric model-fitting for the 17 different combinations of age and SES index. We will here consider their results for ages 7 and older. According to Hanscombe et al.'s Table 2, IQ at ages 2, 3, and 4 correlated significantly, though modestly, with IQ at the later ages—in no case did the correlation exceed 0.30. We regard this as supporting evidence for our assertion that GCA does not emerge as a stable trait until age 4 or 5.

For SES index #1, the AIC-preferred model included only a $C \times SES$ effect at ages 9 and 14, only an $A \times SES$ effect at age 10, and no SES-moderation at ages 7 and 12. For SES index #2, the AIC-preferred model included only a $C \times SES$ effect at ages 9, 10, and 12, and no SES-moderation at ages 7 and 12. For SES index #3, the AIC-preferred model included only a $C \times SES$ effect at ages 9, 10, 12, and 14. Thus, the moderation effect most consistently implicated is moderation of the shared-environmental component. Although the $C \times SES$ effect is in the hypothesized direction in almost all cases, the only clear effect on the additive-genetic component (with SES index #1 at age 10) was in the opposite direction from hypothesis (i.e., decreasing additive-genetic variance with increasing SES). Thus, Hanscombe et al. only replicate the $C \times SES$ effect.

Most recently, Bates et al. (2013) conducted a replication study in a sample of 1,702 twins (in 851 pairs) from the MacArthur Foundation's Midlife in the United States survey. GCA was measured using five specific ability tests, the scores from which were standardized, and then summed within-person, to produce a single composite score. These ability scores were corrected for age, sex, and age squared prior to analyses. Parental SES was operationalized as a combination of parental income and occupational status. In the final biometric model Bates et al. selected, the only biometric moderation effect present was $A \times SES$, and thus they report replication of Turkheimer et al.'s (2003) $A \times SES$ result.

Supplementary References

Asbury, K., Wachs, T. D., & Plomin, R. (2005). Environmental moderators of genetic influence on verbal and nonverbal abilities in early childhood. *Intelligence, 33,* 643-661. doi:10.1016/j.intell.2005.03.008

Kremen, W. S., Jacobson, K. C., Xian, H., Eisen, S. A., Waterman, B., Toomey, R., … Lyons, M. J. (2005). Heritability of word recognition in middle-aged men varies as a function of parental education. *Behavior Genetics, 35*(4), 417-433. doi:10.1007/s10519-004-3876-2

Miller, M. B., Basu, S., Cunningham, J., Eskin, E., Malone, S. M., Oetting, W. S., … McGue, M. (2012). The Minnesota Center for Twin and Family Research Genome-Wide Association Study. *Twin Research & Human Genetics, 15*(6), 767-774.

Sattler JM. (1974). *Assessment of Children (Revised).* Philadelphia: W. B. Saunders Company.

Tucker-Drob, E. M., & Harden, K. P. (2012). Learning motivation mediates gene-by-socioeconomic status interaction on mathematics achievement in early childhood. *Learning and Individual Differences, 22,* 37-45.

Tucker-Drob, E. M., Rhemtulla, M., Harden, K. P., Turkheimer, E., & Fask, D. (2011). Emergence of a gene × socioeconomic status interaction on infant mental ability between 10 months and 2 years. *Psychological Science, 22*(1), 125-133. doi:10.1177/0956797610392926

Wilson, R. S. (1983). The Louisville Twin Study: Developmental synchronies in behavior. *Child Development, 54*, 298-316.

Table S1. Descriptive characteristics of the primary sample.

	MZ twins, Older cohort	MZ twins, Younger cohort	DZ twins, Older cohort	DZ twins, Younger cohort	Ado-Ado Sibs	Bio-Bio Sibs	Mixed Sibs
N #families	832 416	1571 789	419 210	929 469	567 285	415 208	242 122
Female(%)	54.46	50.10	52.98	52.64	57.67	51.33	54.13
Age at Intake M SD	17.46 0.47	11.78 0.44	17.51 0.44	11.79 0.41	14.82 2.07	15.05 1.75	14.94 1.87
Mean FSIQ	100.03	103.23	99.16	103.87	106.55	107.60	108.29

Table notes: “Ado-Ado” = both siblings adopted, “Bio-Bio” = both siblings are biological offspring of parents, “Mixed” = one sibling is adopted and one is biological offspring.

Table S2. Parameter Estimates and Standard Errors from Two Best-Approximating Models #15 and #17

Parameter	Model #15		Model #17		Model-average: Estimate (SE)
	Estimate (SE)	Standardized Estimate	Estimate (SE)	Standardized Estimate	
Intercept, Twins	102.05 (1.43)	-0.04	102.06 (1.43)	-0.04	--
Intercept, Biological SIBS offspring	106.15 (1.72)	0.25	106.17 (1.72)	0.25	--
Intercept, Adoptees	110.19 (2.62)	0.16	110.25 (2.62)	0.16	--
Age, main effect	-0.47 (0.09)	-0.09	-0.47 (0.09)	-0.09	--
Sex, main effect	-3.91 (0.44)	-0.14	-3.92 (0.44)	-0.14	--
SES, main effect, adoptees	7.07 (2.80)	0.12	7.07 (2.80)	0.12	6.96 (2.73)
SES, main effect, bio offspring	16.11 (1.10)	0.28	16.10 (1.10)	0.28	16.05 (1.10)
Main effect of A	8.68 (0.62)	0.73	8.43 (0.60)	0.73	--
Main effect of C	4.93 (0.64)	0.35	4.93 (0.64)	0.35	--
Main effect of E	6.12 (0.32)	0.47	6.57 (0.13)	0.47	--
$A \times SES$ effect	2.79 (0.84)	0.05	3.26 (0.78)	0.06	2.97 (0.96)
$E \times SES$ effect	0.83 (0.58)	0.01	0 (fixed)	0	0.89 (0.58)

Notes: Model #15 had the smallest AICc, and Model #17 had the second smallest. For the sake of side-by-side comparison, the rightmost column presents model-averaged results for select parameters of interest. Standardized estimates are the resulting point estimates when all variables have been standardized to zero mean and unit variance. Model #15 included $A \times SES$ and $E \times SES$ effects, whereas Model #17 only included an $A \times SES$ effect.

Table S3. Parameter estimates from a “full” model with all parameters free.

Parameter	Estimate (SE)
Intercept, Twins	102.27 (1.43)
Intercept, Biological SIBS offspring	106.38 (1.72)
Intercept, adoptees	110.67 (2.60)
Age, main effect	-0.48 (0.09)
Sex, main effect	-3.92 (0.45)
SES, main effect, adoptees	6.77 (2.75)
SES, main effect, bio offspring	15.95 (1.10)
Main effect of A	5.66 (2.87)
Main effect of C	3.49 (6.35)
Main effect of E	6.11 (0.32)
Full-sib genetic correlation	0.43 (0.07)
Twin effects	1.29 (1.16)
$A \times SES$ effect	-1.45 (2.70)
$C \times SES$ effect	23.50 (11.71)
$E \times SES$ effect	0.85 (0.57)
$A \times Age$ effect	0.18 (0.18)
$C \times Age$ effect	0.16 (0.52)
$A \times Age \times SES$ effect	0.30 (0.19)
$C \times Age \times SES$ effect	-1.76 (0.90)

Table notes: All models of Blocks #1, #2, and #3 are nested within this model, which was fitted *post-hoc*.