Intelligence 77 (2019) 101400

Contents lists available at ScienceDirect

Intelligence

journal homepage: www.elsevier.com/locate/intell

Multivariate analysis of the Scarr-Rowe interaction across middle childhood and early adolescence

Evan J. Giangrande[a,*], Christopher R. Beam[b], Sarah Carroll[a], Lucas J. Matthews[a], Deborah W. Davis[c], Deborah Finkel[d], Eric Turkheimer[a]

[a] Department of Psychology, University of Virginia, PO Box 400400, Charlottesville, VA 22904-4400, USA
[b] Department of Psychology, University of Southern California, Los Angeles, CA, USA
[c] Department of Pediatrics, University of Louisville School of Medicine, Louisville, KY, USA
[d] Department of Psychology, Indiana University Southeast, New Albany, IN, USA

ARTICLE INFO

Keywords:
Intelligence
SES
Gene-environment interaction
Multivariate methods

ABSTRACT

Numerous studies have found interactions between socioeconomic status (SES) and the heritability of cognitive ability in samples from the United States, with individuals from lower SES backgrounds showing decreased heritability compared to those reared in higher SES environments. However, nearly all published studies of the Scarr-Rowe interaction have been univariate and cross-sectional. In this study, we sought to maximize statistical power by fitting multivariate models of gene (G) x SES interaction, including longitudinal models. Cognitive ability data collected at up to five time points between ages 7 and 15 years were available for 566 twin pairs from the Louisville Twin Study. We used multilevel and latent factor models to pool intelligence subtest scores cross-sectionally. To examine interactions longitudinally, we fit latent growth curve models to IQ scores. Power analysis results indicated that the multivariate approach substantially boosted power to detect G x SES interaction. The predicted interaction effect was observed at most ages in cross-sectional multivariate analyses. In longitudinal analyses, we found significant G x SES interactions on mean-level (intercept) full scale IQ and performance IQ ($ps < .001$), but not verbal IQ intercept ($p = .08$). SES did not significantly moderate the heritability of change in IQ over time (slope). Interaction appeared to be driven by DZ twin correlations decreasing more substantially as a function of higher SES than MZ correlations.

1. Introduction

Low socioeconomic status (SES) is associated with negative outcomes in a variety of important domains, including cognitive ability (Bradley & Corwyn, 2002). Turkheimer, Haley, Waldron, D'Onofrio, and Gottesman (2003) observed an interaction of SES and the heritability of IQ in 7-year-old U.S. twins, wherein children from lower SES families showed reduced heritability compared to more affluent peers. This finding supported the Scarr-Rowe hypothesis, which holds that environmental disadvantage hinders the ability of individuals reared in lower SES households to realize their intellectual potential (Rowe, Jacobson, & Van den Oord, 1999; Scarr-Salapatek, 1971).

Modification of cognitive performance heritability by SES has since been observed in most studies using U.S. samples, and a recent meta-analysis of such studies found a moderately sized interaction effect (Tucker-Drob & Bates, 2016). Significant gene (G) x SES interaction has been observed across the life span, including in early childhood (Rhemtulla & Tucker-Drob, 2012; Tucker-Drob, Rhemtulla, Harden, Turkheimer, & Fask, 2011), middle childhood (Turkheimer et al., 2003), adolescence (Harden, Turkheimer, & Loehlin, 2007; Rowe et al., 1999), and adulthood (Bates, Lewis, & Weiss, 2013). Several U.S. studies, however, have failed to find significant moderation (Grant et al., 2010; Kremen et al., 2005), including a recent study by Figlio, Freese, Karbownik, and Roth (2017). G x SES interaction is not typically present in samples from Western Europe and Australia, where factors associated with environmental enrichment (e.g., quality education and healthcare) are more widely accessible (Grasby, Coventry, Byrne, & Olson, 2017; Tucker-Drob & Bates, 2016). There are, however, exceptions to this pattern as well, especially in cohorts from previous generations (Fischbein, 1980; Turkheimer, Beam, Sundet, & Tambs, 2017).

Despite the growing body of work on G x SES interaction on cognitive ability, existing studies have been limited in several important ways. First, many studies have lacked sufficient statistical power, decreasing the likelihood of detecting interaction effects (Tucker-Drob &

* Corresponding author.
E-mail address: ejg2br@virginia.edu (E.J. Giangrande).

https://doi.org/10.1016/j.intell.2019.101400
Received 22 July 2018; Received in revised form 13 August 2019; Accepted 27 September 2019
Available online 02 November 2019

Table 1
Demographic and descriptive information.

Age (years)	n MZ/DZ Pairs	n same/opp. sex pairs	n pairs both female/both male/opp. sex	% female	% Caucasian	FSIQ	SES
7	235/236	374/97	204/170/97	53.61	88.85	98.34 (14.06)	47.99 (26.80)
8	250/253	401/102	215/186/102	52.88	90.95	101.80 (14.02)	47.53 (26.38)
9	191/199	297/94	160/137/94	52.94	88.87	102.82 (14.48)	46.84 (27.15)
12*	71/82	113/40	55/58/40	49.02	81.37	100.76 (14.49)	44.75 (29.11)
15	191/184	304/71	164/140/71	53.20	93.07	99.87 (14.02)	46.41 (26.20)
All Ages	282/284	446/120	236/210/120	52.30	90.37	100.68 (14.25)	47.80 (26.56)

* Cross-sectional analyses were not performed on age 12 data due to insufficient sample size. Opp: opposite. FSIQ and SES presented as mean (standard deviation).

Table 2
Correlations of full scale IQ across ages and SES.

	IQ 7	IQ 8	IQ 9	IQ 12	IQ 15	SES
IQ 7	1	–	–	–	–	–
IQ 8	.88	1	–	–	–	–
IQ 9	.88	.90	1	–	–	–
IQ 12	.85	.88	.88	1	–	–
IQ 15	.78	.82	.83	.91	1	–
SES	.39	.36	.37	.50	.36	1

Pearson's correlation coefficients. All pairwise correlations were significant ($p < .05$).

Table 3
Missing cognitive data information for longitudinal analyses.

Age (years)	7	8	9	12	15
7	0.83	–	–	–	–
8	0.74	0.89	–	–	–
9	0.55	0.66	0.69	–	–
12	0.27	0.25	0.25	0.27	–
15	0.54	0.64	0.51	0.19	0.66

Values on the diagonal indicate the proportion of the total sample that had cognitive data at each age. Off-diagonal values represent the proportion of the total sample available to calculate a covariance between cognitive measures at two ages.

Bates, 2016). Second, few studies have examined G x SES interaction longitudinally. Tucker-Drob et al. (2011) observed significant interaction on change in mental ability between ages 10 months and 2 years. Rhemtulla and Tucker-Drob (2012) found that individual differences in mathematics skills (but not reading) among four-year-olds were moderated by SES, and that this interaction was not explained by interaction effects on mental ability at age 2. However, Rhemtulla and Tucker-Drob (2012) did not test for interaction effects on change in cognitive ability between 2 and 4 years. We are unaware of previous studies that have examined G x SES interaction longitudinally at later ages. Thus, the extent to which SES may affect the heritability of *change* in cognitive ability over time (in addition to performance at a given age) has been largely unstudied.

In the present study, we sought to address these limitations by investigating heritability x SES interaction across middle childhood and early adolescence using cross-sectional and longitudinal multivariate techniques. Compared to univariate twin models, both cross-sectional and longitudinal multivariate models offer increased power, as long as observed measures are correlated (Schmitz, Cherny, & Fulker, 1998). Using multiple observations increases the amount of information available for each subject, thereby improving measurement reliability and reducing error (Allison, Allison, Faith, Paultre, & Pi-Sunyer, 1997). Put differently, the multivariate approach increases effect sizes (and therefore power) by decreasing within-group variance. In theory, pooling multiple measures of cognitive ability should increase power to detect G x SES interaction effects, helping address the power limitations that have plagued previous studies. However, previous studies have not demonstrated this potential power boost empirically. We addressed this gap by comparing three multivariate models of G x SES interaction (two cross-sectional, one longitudinal) to a more traditional univariate model, and to each other. Along with its possible power benefits, the longitudinal model also enabled us to examine whether SES moderates the heritability of change in cognitive performance over late childhood/early adolescence.

Data were drawn from the recently revived Louisville Twin Study (LTS; Rhea, 2015; Wilson, 1983). Although the LTS generated one of the most comprehensive data sets on the early cognitive development of U.S. twins ever collected (Rhea, 2015), G x SES interaction has not been thoroughly explored in that sample. A preliminary univariate study observed G x SES interaction in 7-year-old LTS twins that approached but did not reach statistical significance ($p < .07$; Turkheimer, Beam, & Davis, 2015). Since that report, additional cognitive data for other ages (up to 15 years) have been recovered, increasing the overall sample size by approximately 100 twin pairs. Furthermore, although preliminary analyses used index-level cognitive performance scores (i.e., full scale IQ, performance IQ, and verbal IQ), subtest scores are also available for all LTS twins at all measurement occasions, making it possible to conduct multivariate analyses of the common variance across subtests. Finally, a large subset of LTS twins participated in cognitive testing at multiple time points, enabling us to perform longitudinal analyses of cognitive performance heritability x SES interaction across middle childhood and early adolescence.

By developing multivariate models and applying them to LTS data, we intended to 1) investigate the Scarr-Rowe interaction comprehensively in that important sample, substantially expanding upon the previous preliminary report, and 2) demonstrate the utility and robustness of those models, particularly in comparison to traditional univariate techniques. We hypothesized that we would observe modest interaction effects at all ages; specifically, we expected the proportion of variance in cognitive performance attributable to additive genetic factors (A) and shared environmental factors (C) to increase and decrease, respectively, as a function of SES. We also predicted that multivariate models would be more statistically powerful than univariate analyses.

2. Method

2.1. Participants

Data collection for the LTS ran from 1957 until the late 1990s (Rhea, 2015; Wilson, 1983). Participants were all from the Louisville, Kentucky area. Twin pairs were included in the current analyses if 1) both twins in a set participated in at least one cognitive assessment at ages 7, 8, 9, 12, or 15; and 2) family-level SES was available. We analyzed data from 566 twin pairs in total (Table 1; 282 monozygotic (MZ), 284 dizygotic (DZ); 236 same sex female, 210 same sex male, 120 opposite sex). Zygosity was determined by blood serum analysis. The sample was of average intelligence and SES (Table 1; Table 2) and 90.37% Caucasian. Age 12 data were omitted from cross-sectional analyses due to insufficient sample size. 80.04% of the sample participated in data collection at three or more ages. Missing data information for longitudinal analyses is presented in Table 3.

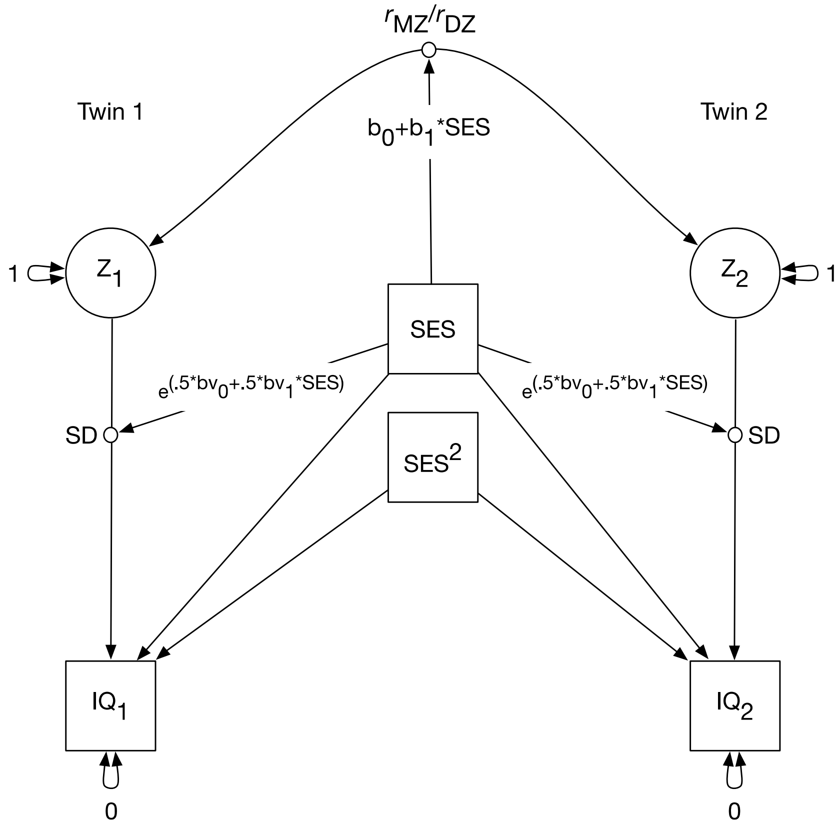

Fig. 1. Modified twin correlation model.
IQ: placeholder for the observed cognitive variables we analyzed. Z: latent variable that standardized the cognitive variable to a mean of 0 and standard deviation of 1, thereby transforming the twin covariances into correlations. r_{MZ}/r_{DZ}: monozygotic/dizygotic twin correlations for cognitive ability. SD: standard deviation of the cognitive variable. We fit separate linear models of SES ($b_0 + b_1 * SES$) to rMZ and rDZ to examine whether twin correlations changed as a function of SES. A log-linear model of SES ($e^{(0.5*bv0+\ 0.5*bv1\ *\ SES)}$) was fit to the phenotypic variance to account for phenotypic heteroscedasticity. The 0.5 term in the exponential expression for the variance is included because SD is a standard deviation, not a variance.

2.2. Measures

Three versions of the Wechsler Intelligence Scale for Children (WISC) were administered during the LTS: the WISC, WISC-R, and WISC-III (Wechsler, 1949, 1974, 1991). We used age-scaled index and subtest WISC scores in cross-sectional analyses, and only index scores in longitudinal analyses. SES at initial registration in the study was measured with the Hollingshead Four Factor Index of Socioeconomic Status, which is a continuous zero to 100-point scale based on parental occupation, education, sex, and marital status (Hollingshead, 1975). SES was normally distributed in our sample (Q1, median, and Q3 = 24, 49, and 70, respectively).

2.3. Procedure

We used R to calculate descriptive statistics and prepare the data (R Core Team, 2018). Twin models were fit in Mplus Version 8 (Muthén & Muthén, 2017) using full information maximum likelihood estimation to handle missing data.

2.3.1. Univariate analyses

To build upon existing univariate, cross-sectional examinations of G x SES interaction, and to provide a baseline against which to compare our multivariate models, we first modeled MZ and DZ covariances for each index and subtest score as a function of standardized SES at ages 7, 8, 9, and 15. We used a modified twin correlation model (MTCM; Fig. 1; Turkheimer et al., 2017), which differs from the commonly used Purcell model (Purcell, 2002) in several important ways. First, cognitive variables are standardized *within* the MTCM, meaning that the twin covariances are correlations. This is done by creating a latent variable (Z) that has a variance of one and is indicated by the observed cognitive measure (e.g., IQ, as depicted in the figure), which has its residual variance fixed to zero. The internal standardization results in a factor loading weight equal to the observed standard deviation of the phenotype (SD), which can then be examined for heteroscedasticity with respect to the moderator using an exponential function.

Second, in the MTCM, SES linearly modifies the MZ and DZ twin correlations (r_{MZ}/r_{DZ}). The twin correlations and their moderation can then be linearly transformed into additive genetic (A), shared environmental (C), and non-shared environmental (E) variance components. This contrasts the Purcell model, wherein SES modifies the paths from the ACE components to the measured outcome. The Purcell model is therefore implicitly a quadratic model of the ACE variances, which are necessarily constrained to be greater than zero. The MTCM's focus on moderation of the twin correlations allows the correlations to assume moderated values that would result in negative C estimates, which violates the ACE parameterization of the classical twin model. Permitting the ACE parameters to be modeled as negative makes it possible to model the twin correlations accurately, particularly when the DZ twin correlation is less than half of the MZ correlation, as has been observed in previous studies of cognitive ability (Turkheimer et al., 2017). Finally, in addition to controlling for linear main effects of SES on cognitive ability as in the Purcell model, we also controlled for quadratic main effects in univariate analyses.

Consistent with the classical twin model, we constrained the means and variances of cognitive measures to be equal across twins in a pair.

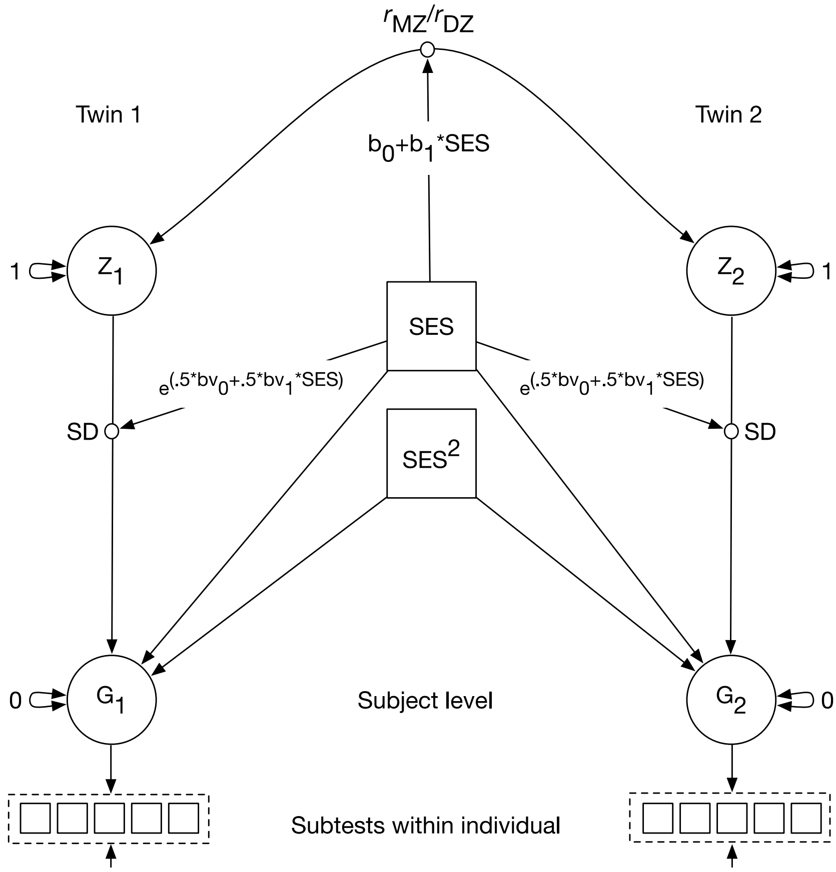

Fig. 2. Multilevel model.
Empty squares represent the 12 WISC subtests. G: general cognitive ability. The top part of the model is identical to the modified twin correlation model (MTCM) presented in Fig. 1. For each participant, we fit the MTCM to the 12 subtests simultaneously and adjusted the standard errors.

Expected MZ and DZ covariances were 1 and 0.5, respectively. We tested for significant G x SES interaction (in both univariate analyses and the multivariate models discussed below) using a Wald test with two degrees of freedom, which examined whether the A and/or C moderation parameters differed significantly from zero as a function of linear SES.

2.3.2. Cross-sectional multivariate analyses

Within each age, we pooled information from all 12 WISC subtests using two multivariate models, both of which were extensions of the univariate model described above. The first was a multilevel model (Fig. 2). The top part of the model is the same as the univariate MTCM described above. The 12 observed subtest scores (represented as empty squares) were treated as a repeated measure nested within a subject-level cognitive ability latent variable (G). The MTCM was fit to all available subtest scores simultaneously within each individual, and standard errors were corrected to account for the fact that the subtest scores came from the same individual.

The second was a latent factor model (Fig. 3) in which the MTCM was applied to a common factor (G) that was estimated from the 12 WISC subtests (depicted as empty squares) for each twin. We fixed the loading of the first subtest to one, thereby fixing the variance of the latent factor, and then standardized the latent factor inside the model using the same method as described for the univariate analyses. Residual variances of the subtest scores were correlated across twins. SES modified the MZ and DZ twin correlations for latent cognitive ability in the same manner as in the univariate MTCM.

The multilevel model and the latent factor model differed in that the former analyzed both common and unique variance across multiple subtests for each participant, whereas the latter only analyzed common variance. Because of this, the multilevel model was expected to result in smaller twin correlations than the latent factor model. More theoretically, these models represent two ways of handling subtest scores. The multilevel model treats subtest scores as multiple observations of a participant's ability. The standard errors of the parameter estimates are corrected to take subject-level covariation among the subtests into account. The latent factor model, in contrast, treats each subtest score as a manifestation of a single underlying ability.

2.3.3. Longitudinal multivariate analyses

Next, we examined G x SES interaction longitudinally by fitting a latent growth curve (LGC) model to full scale, performance, and verbal IQ (FSIQ, PIQ, and VIQ, respectively) data from ages 7, 8, 9, 12, and 15 (Fig. 4). As with the cross-sectional multivariate models, our LGC model is largely an extension of the univariate MTCM. In LGC analyses, individual differences in phenotypic change are modeled as random effects using two factors. The intercept factor uses information from all available observations to estimate performance at the first age of

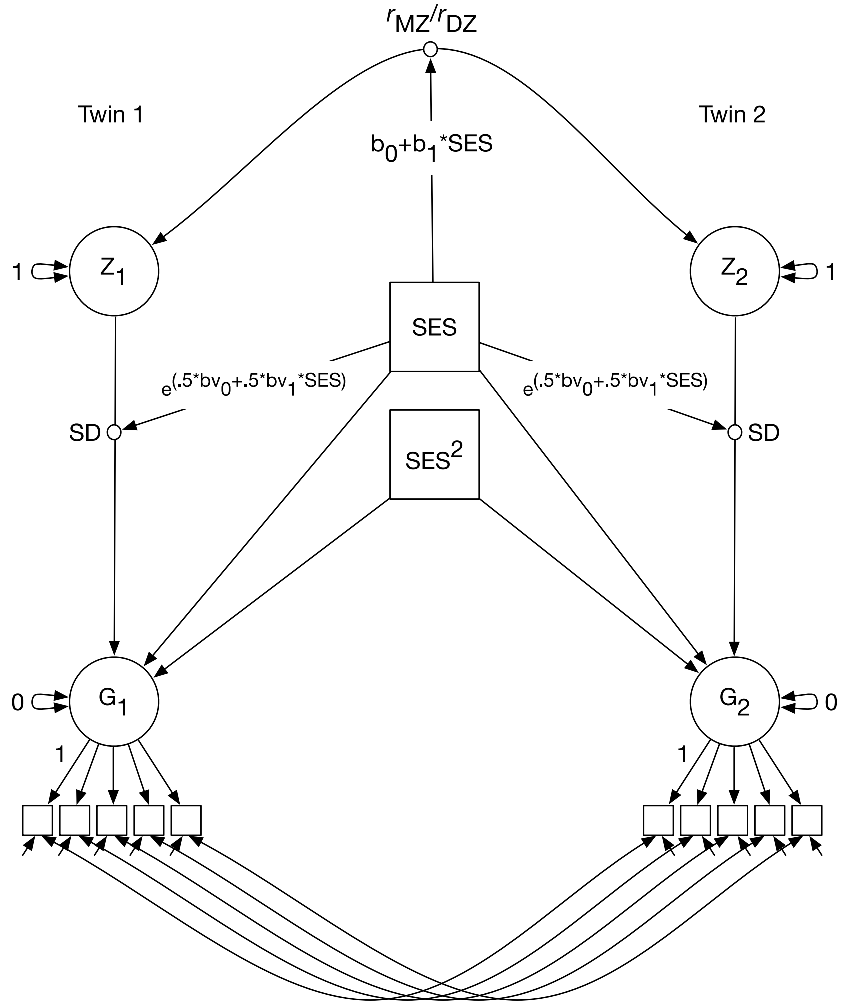

Fig. 3. Latent factor model.
G: latent cognitive factors generated from the 12 WISC subtests, which are represented by empty squares. The top part of the model is identical to the modified twin correlation model (MTCM) presented in Fig. 1. We fit the MTCM to each individual's latent factor.

measurement, while the slope factor indexes rate of change from that initial performance over time. In this study, we created latent IQ intercept (I) and slope (S) factors for each twin and fixed the intercept at the first time point (7 years). IQ loadings on the intercept factor were all fixed to 1, while slope loadings were weighted to model the time elapsed between observations. SES modified the intercept and slope twin correlations directly, controlling for linear effects of SES. As in the univariate MTCM, the modification parameters were linearly transformed into equivalent values of the A and C components. Separate Wald tests were performed for the intercept and slope factors. Thus, this model tested whether SES modified 1) the heritability of IQ at the first time point (intercept) and 2) the heritability of change in IQ over time (slope). To model autoregressive effects (i.e., the extent to which the variance of one observation explained the variance of subsequent observations), observed IQ at each age was regressed on IQ measured at the previous age. IQ scores had residual variances (E), which correlated across twins for corresponding measurement occasions.

2.3.4. *Power analyses*

To test whether our multivariate models were more statistically powerful than the univariate MTCM, as hypothesized, we performed power analyses of all four models. We simulated 1000 data sets that each included 250 MZ and 250 DZ twins, roughly equivalent to our sample sizes at each age. The structure of the covariance across subtests, both within and between twins, was specified to reflect the covariance structure observed in the LTS data. SES values were randomly drawn from a uniform distribution with a mean of zero and standard deviation of one. MZ and DZ correlations in each data set were calculated as 0.6–0.01 * SES and 0.4–0.025 * SES, respectively. Thus, as SES increased, the DZ correlation decreased more than the MZ correlation. We then fit our univariate, multilevel, and latent factor models to the 1000 data sets. The univariate power analysis was run on simulated data from a single, randomly-selected subtest. Power analyses of the cross-sectional multivariate models used data from all 12 subtests.

The power analysis of the LGC model followed an analogous procedure. We generated an additional 1000 data sets (i.e., independent of the data sets created for the univariate and cross-sectional multivariate

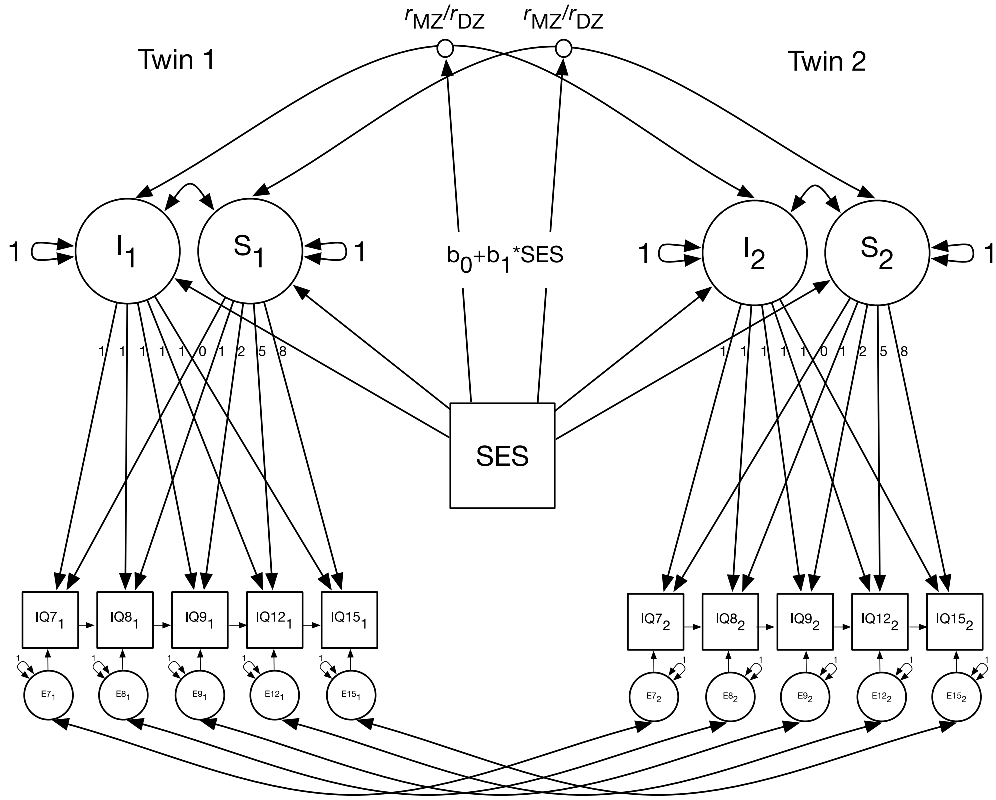

Fig. 4. Latent growth curve model.
I: latent intercept factor. S: latent slope factor. E: residual variance. r_{MZ}/r_{DZ}: monozygotic/dizygotic twin correlations for cognitive ability. For both I and S, we fit separate linear models of SES ($b_0 + b_1$ * SES) to *r*MZ and *r*DZ to examine whether twin correlations changed as a function of SES.

power analyses), which had a covariance structure comparable to that observed in the longitudinal data. We analyzed the LGC model's power to detect a significant interaction effect on the intercept.

3. Results

3.1. Power analyses

Our multivariate models offered substantially more power to detect significant G x SES interaction than the univariate MTCM (Fig. 5a). However, multivariate analyses were still underpowered compared to an 80% threshold, and results should be considered in light of this fact. Specifically, lower power decreases confidence in observed null G x SES effects. It also enlarges standard errors, increasing variability in the observed magnitude of interaction effects, particularly across ages in cross-sectional analyses. In the power analysis of the univariate MTCM, the Wald test was significant in only 105 out of 994 converged models, indicating that power was only 10.56%. This was less than half of the power in the multilevel model, which equaled 22.30% (all models converged, with 223 yielding a significant Wald test). Power in the latent factor model was 60.98% (558 out of 915 Wald tests were significant). In the LGC model, power to detect significant interaction on the intercept was 37.00% (356 out of 962 Wald tests were significant).

3.2. Univariate analyses

In univariate analyses, significance of the Wald test fluctuated across ages and cognitive measures in a manner that did not follow a discernible pattern (Fig. 5b), likely because of power limitations. Results provided tentative evidence of G x SES interaction: modified twin correlations tended to assume values that resulted in A increasing as a function of SES while C decreased (Supplementary table 1; mean b_1 A = 0.03, 57% positive; mean b_1 C = −0.06; 72% negative). Overall, the univariate results suggested that increasing power using a multivariate approach may be worthwhile.

3.3. Cross-sectional multivariate analyses

Complete multilevel and latent factor model results are presented in Table 4 and Fig. 6.

3.3.1. Multilevel model

We did not observe significant G x SES interaction in multilevel analysis of age 7 data ($p > .05$). However, SES modified twin correlations for cognitive ability at ages 8 and 9 such that DZ correlations decreased more than MZ correlations as a function of higher SES. When twin correlations were transformed into ACE variances, the predicted pattern of A increasing and C decreasing as a function of SES was observed. Wald tests of the b_1 A and C parameters were significant at both ages 8 and 9 (age 8: χ^2 (2, n = 503 pairs) = 11.12, p = .004; age 9: χ^2

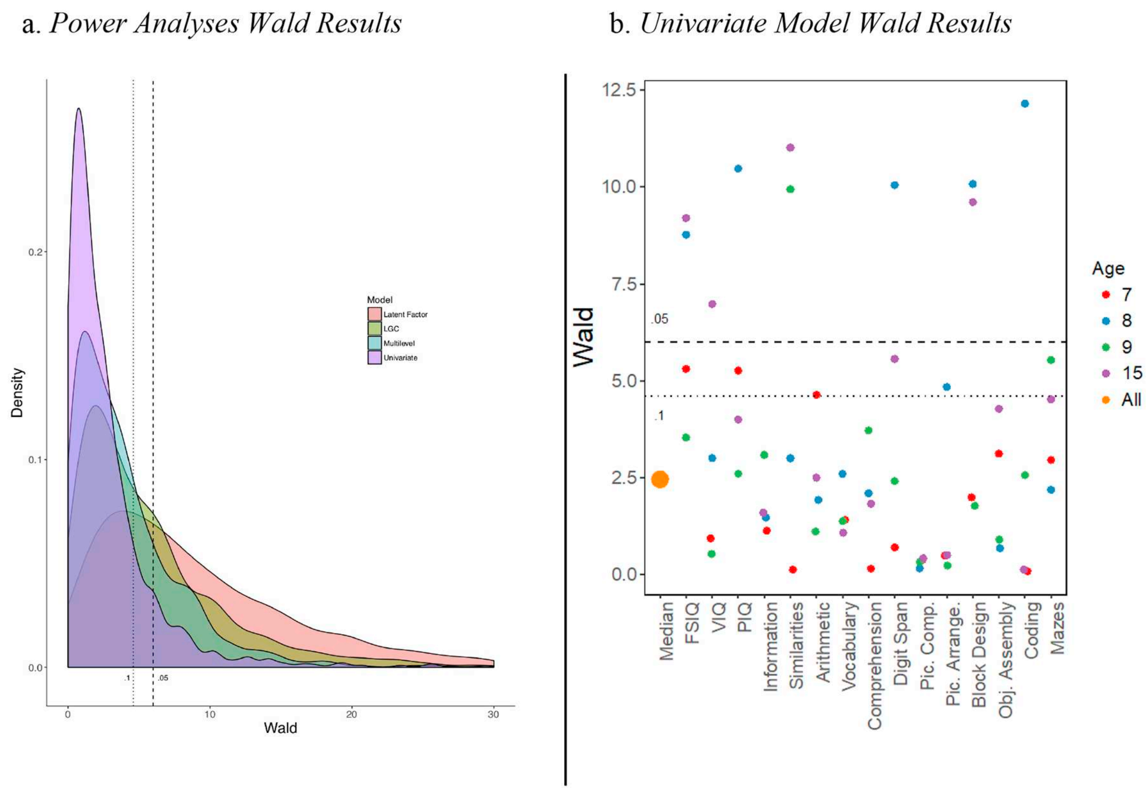

Fig. 5. (a) Power analyses Wald results. (b) Univariate model Wald results.
5a) Density plot of Wald chi-square results from power analyses. X axis upper limit set at 30 for ease of visualization. Dotted and hashed lines: significance cutoffs of $p = .1$ and $p = .05$, respectively. 5b) Points above dotted and hashed lines: significant at $p < .1$ and $p < .05$, respectively. Values < 2 are jittered slightly. Orange dot: overall median Wald for all measures and ages. Pic: picture. Comp: comprehension. Arrange: arrangement. Obj: object. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(2, n = 390 pairs) = 9.18, p = .01). Although the Wald test was also significant at age 15 (χ^2 (2, n = 375 pairs) = 12.15, p = .002), the b_1 A estimate was slightly negative, which did not conform to the predictions of our model.

3.3.2. Latent factor model

As predicted, the latent factor model yielded larger twin correlations than the multilevel model. Results of latent factor analyses suggested that 7-, 8-, and 9-year-old DZ twins diverged in cognitive ability as a function of higher SES more than MZ twins. As a result, A increased and C decreased as a function of SES at all three ages. The Wald test of the interaction was significant at age 7 (χ^2 (2, n = 471 pairs) = 6.25, p = .04), but not age 8 or 9 ($ps > .05$). Although age 15 twin correlations and corresponding A and C variances did not follow the expected interaction pattern, neither A nor C was significantly different from zero ($p > .05$).

3.4. Longitudinal multivariate analyses

Results of LGC models are presented in Table 5, Fig. 7, and Supplementary tables 2 and 3. We observed significant SES modification of cognitive performance intercept twin correlations for FSIQ (χ^2 (2, N = 566 pairs) = 17.02, $p < .001$) and PIQ (χ^2 (2, N = 566 pairs) = 14.22, $p < .001$). As in most of the cross-sectional multivariate results, DZ correlations decreased more than MZ correlations as a function of higher SES. This drove the interaction observed after transformation into ACE variances, where A increased and C decreased as a function of SES. SES did not significantly modify VIQ intercept twin correlations, or slope correlations for FSIQ, PIQ, or VIQ ($ps > .05$).

4. Discussion

In this study, we conducted a comprehensive examination of G x SES interaction of cognitive ability in LTS data using novel multivariate models. Cross-sectional and longitudinal multivariate models were considerably more statistically powerful than our univariate model, as hypothesized, but multivariate analyses were still somewhat underpowered. Collectively, results provide further evidence of G x SES interaction on cognitive ability across middle childhood and early adolescence among U.S. twins. In 7 of 11 multivariate analyses, SES modified twin correlations for cognitive ability such that individuals from more affluent families showed increased heritability compared to less privileged peers. Significant interaction effects were observed in both cross-sectional and longitudinal multivariate analyses, and all but one were in the expected direction (i.e., A increasing and C decreasing as a function of higher SES). Consistent with the results of Tucker-Drob and Bates' (2016) meta-analysis, which reported an interaction effect of 0.074 for U.S. G x SES studies, we observed effect sizes in a similar range. Our analyses provided a thorough documentation of the Scarr-Rowe interaction in the LTS, an important study of twin development

Table 4
Cross-sectional multivariate results.

Model	Age	Wald	Linear ME	Quad ME	b0 rMZ	b1 rMZ	b0 rDZ	b1 rDZ	b0 A	b1 A	b0 C	b1 C	b0 E	b1 E
Multilevel	7	4.78	0.24 (0.03)*	−0.07 (0.03)*	0.53 (0.01)*	−0.03 (0.01)*	0.38 (0.03)*	−0.02 (0.03)	0.30 (0.06)*	−0.01 (0.06)	0.24 (0.05)*	−0.02 (0.05)	0.47 (0.01)*	0.03 (0.01)*
	8	11.12*	0.22 (0.02)*	−0.03 (0.03)	0.57 (0.01)*	−0.03 (0.01)*	0.42 (0.02)*	−0.06 (0.02)*	0.31 (0.05)*	0.06 (0.05)	0.26 (0.05)*	−0.09 (0.04)*	0.43 (0.01)*	0.03 (0.01)*
	9	9.18*	0.23 (0.03)*	−0.02 (0.03)	0.61 (0.02)*	−0.03 (0.02)	0.38 (0.03)*	−0.07 (0.03)*	0.45 (0.06)*	0.08 (0.06)	0.16 (0.05)*	−0.11 (0.05)*	0.39 (0.02)*	0.03 (0.02)
	15	12.15*	0.22 (0.03)*	−0.02 (0.03)	0.62 (0.02)*	−0.05 (0.02)*	0.37 (0.03)*	−0.04 (0.03)	0.49 (0.06)*	−0.02 (0.06)	0.13 (0.06)*	−0.04 (0.05)	0.39 (0.02)*	0.05 (0.02)*
Latent	7	6.25*	0.96 (0.10)*	−0.23 (0.11)*	0.92 (0.02)*	−0.03 (0.02)	0.56 (0.05)*	−0.09 (0.05)	0.71 (0.10)*	0.11 (0.10)	0.21 (0.10)*	−0.14 (0.09)	0.08 (0.02)*	0.03 (0.02)
	8	4.91	0.99 (0.10)*	−0.08 (0.11)	0.91 (0.02)*	−0.02 (0.01)	0.60 (0.05)*	−0.07 (0.04)	0.63 (0.10)*	0.10 (0.09)	0.28 (0.09)*	−0.12 (0.09)	0.09 (0.02)*	0.02 (0.01)
	9	3.31	1.06 (0.12)*	−0.09 (0.13)	0.93 (0.02)*	−0.02 (0.02)	0.58 (0.05)*	−0.07 (0.05)	0.71 (0.11)*	0.09 (0.10)	0.22 (0.10)*	−0.11 (0.10)	0.07 (0.02)*	0.02 (0.02)
	15	4.16	1.02 (0.12)*	−0.03 (0.12)	0.94 (0.01)*	−0.03 (0.04)*	0.60 (0.05)*	−0.01 (0.05)	0.69 (0.11)*	−0.04 (0.10)	0.25 (0.11)*	0.01 (0.10)	0.06 (0.01)*	0.03 (0.01)*

Wald: result of Wald chi-square test with two degrees of freedom. ME: main effect. Quad: quadratic. Parameter estimates presented as value (standard error).

* $p < .05$.

that had only been preliminarily analyzed for G x SES interaction at one age prior to this study (Turkheimer et al., 2015).

Algebraically, G x SES interaction on cognitive ability can arise from MZ twin correlations increasing more rapidly with rising SES than DZ twin correlations, DZ twin correlations decreasing quicker than MZ correlations as SES increases, or a combination of those two mechanisms. The interaction effects that we observed were driven primarily by DZ twin correlations decreasing more substantially than MZ correlations as a function of greater SES; there was little evidence of greater phenotypic convergence in MZ twins at higher levels of SES. When twin correlations were transformed into ACE variance components, greater DZ divergence resulted in A increasing more quickly or decreasing less quickly with rising SES than C. In some analyses, greater divergence in DZ twins resulted in C approaching zero or even taking negative values at higher levels of SES. One possible explanation of this finding could relate to the effects of phenotype-environment correlation (Beam & Turkheimer, 2013); as SES rises, twins are able to self-select into increasingly different environments. Because they are less genetically, and therefore phenotypically, similar, DZ twins select into more discrepant environments than MZ twins. Greater environmental disparity, in turn, causes DZ twins to exhibit larger within-pair phenotypic differences than MZ twins, creating a reciprocal feedback loop between phenotype and environment. Ultimately, this process results in DZ twins being less correlated for cognitive ability than MZ twins as a function of increasing SES, driving G x SES interaction. However, we did not test this hypothesis in the current study, and there are likely other potential explanations in addition to phenotype-environment correlation. Future studies should work to identify the specific mechanisms that underlie the interaction. Enabling twin correlations to assume values that result in negative ACE components, as we did here using the MTCM, may be an important step towards understanding those mechanisms.

The multivariate models developed in this study serve as a significant addition to the existing literature on the Scarr-Rowe interaction, which has previously been examined almost exclusively with univariate, cross-sectional models. Power analysis results indicated that multivariate models of G x SES interaction were superior to more traditional univariate methods because they offer a substantial increase in statistical power. Compared to our univariate model, our multivariate models showed as much as a sixfold increase in power. Even the least powerful multivariate model (multilevel) still was twice as powerful as the univariate MTCM. This is the first study of which we are aware to demonstrate that multivariate methods increase power to detect G x SES interaction on cognitive ability. When thinking about how to boost power, researchers tend to focus on enlarging sample sizes. While that is one solution, recruiting additional participants can be a difficult endeavor when funding is limited and/or data collection has ceased. Therefore, when designing future studies of the Scarr-Rowe interaction, it may be advantageous for researchers to plan on collecting more data per available participant. Using multivariate approaches to re-analyze data sets in which G x SES interaction has previously been examined with univariate models may also be worthwhile.

Although the multivariate approach boosted power considerably, our multivariate models were still underpowered compared to desirable thresholds (e.g., 80%). This is likely due to the sizes of our simulated samples; while these were consistent with LTS sample sizes and not atypical for G x SES interaction studies, larger samples may have enabled power to surpass those thresholds. Power analysis results also are affected by model fit. For example, the power of the LGC model depended in part on how well the model fit the longitudinal pattern of cognitive ability in the simulated data. Future work will be needed to determine the extent to which the power of the multivariate G x SES interaction models is influenced by various characteristics of the data being analyzed. In the current study, power limitations likely decreased our ability to detect significant interaction effects. The null effects we observed should therefore be interpreted cautiously. Power limitations

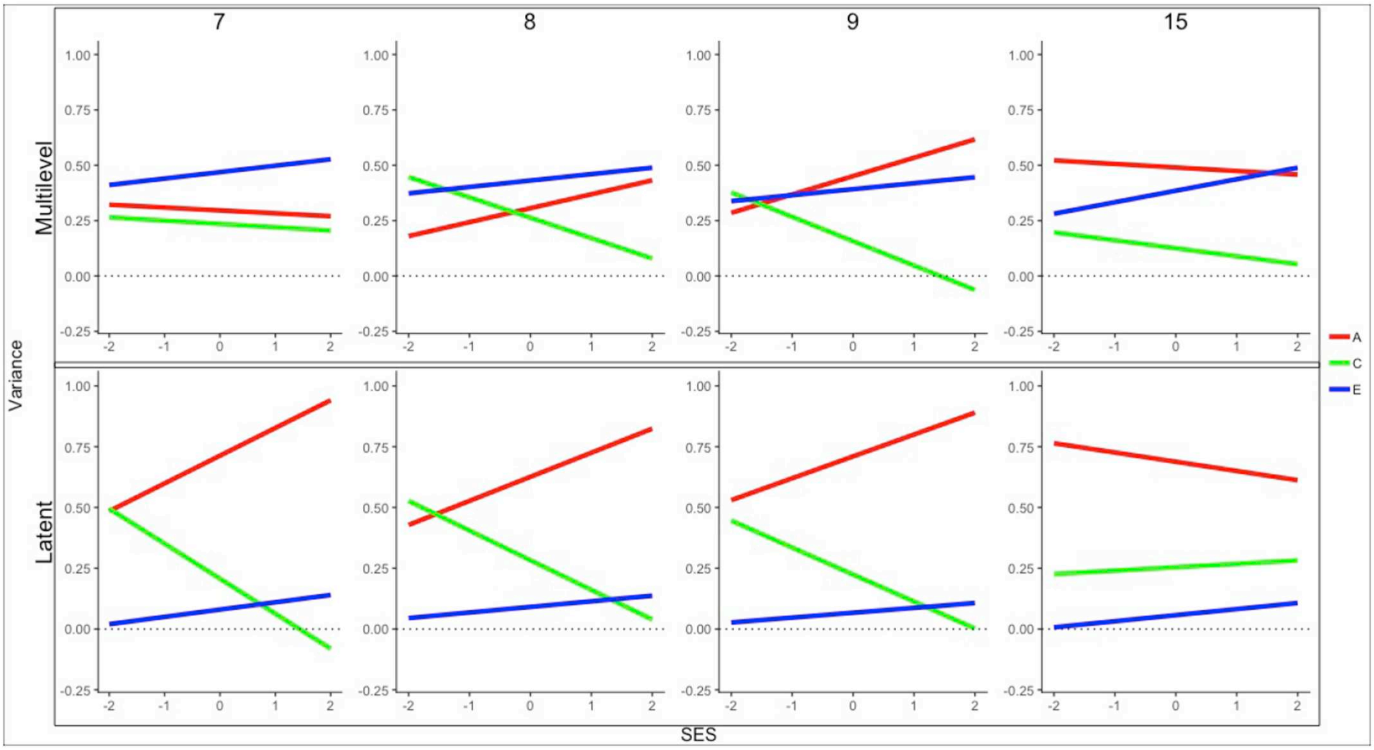

Fig. 6. Cross-sectional multivariate ACE results.Multilevel: results of multilevel cross-sectional model. Latent: results of latent factor cross-sectional model. Variance: proportion of variance in cognitive ability attributable to A, C, and E (presented in red, green, and blue, respectively). SES is standardized to a mean of 0 and standard deviation of 1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

also likely resulted in larger standard errors. This may partially explain why the strength of observed G x SES interaction effects fluctuated across age in cross-sectional analyses.

Trends of the moderated twin correlations and ACE components were generally comparable across the cross-sectional multivariate approaches, although the multilevel model resulted in more significant effects than the latent factor model. At ages 8 and 9 in multilevel analyses and 7, 8, and 9 in latent factor analyses, moderated DZ correlations were more negative than MZ correlations, resulting in A increasing and C decreasing as a function of SES. Interaction effects were significant at ages 8 and 9 in multilevel analyses and at age 7 in latent factor analyses. Age 15 results from both models did not correspond to the expected pattern—moderated MZ correlations were estimated to be slightly more negative than DZ correlations—although the Wald test was only significant in the multilevel model. Given that the age 15 sample was the smallest of the four ages included in cross-sectional analyses, that power was still less than ideal in our multivariate models, and that previous studies have observed G x SES interaction at age 15, it is unlikely that this unexpected result is robust.

The fact that the multilevel model yielded more significant effects than the latent factor model likely stems from substantial differences in the variance analyzed in each model; the former analyzed both common and unique variance across an individual's subtest scores while the latter analyzed only common variance. This resulted in different estimates for twin correlations and corresponding ACE variance components. Given that the Wald test was performed on those components, it is perhaps unsurprising that we observed different levels of statistical significance across the models. Interestingly, more significant effects were observed using the multilevel model even though it is less statistically powerful than the latent factor model. This could underscore the importance of including unique variance in multivariate G x SES interaction models; if G x SES interaction occurs primarily on the unique variance in cognitive ability, choosing not to model the unique variance (as in the latent factor model) could have made it difficult to observe significant interaction effects, even with increased power. More work will be needed to resolve this question.

Importantly, decisions about which models to use should not be made based solely on statistical significance or power, and differences in the number of significant results observed using the multilevel and latent factor models do not necessarily indicate that one is preferable to the other. Given this, it may be acceptable to base the choice between the models on theory. The multilevel model could be favorable in applications where subtests are regarded as multiple repeated observations of cognitive ability, each treated in the same way, whereas the latent factor model may be more appropriate when subtests are treated as indicators of a unitary latent ability and have individual loadings on a general factor. In terms of the variance being analyzed, the multilevel model may be more appropriate when unique variance is thought to play an important role, while the power boost offered by the latent factor model may make it preferable to a researcher focusing on common variance. What is clear is that both the multilevel and latent factor models are preferable to traditional univariate models.

To our knowledge, this was the first study to investigate G x SES interaction on cognitive ability using a latent growth curve model, the

Table 5
Latent growth curve model results.

Index	Factor	Wald	Linear ME	b0 rMZ	b1 rMZ	b0 rDZ	b1 rDZ	b0 A	b1 A	b0 C	b1 C	b0 E	b1 E
FSIQ	Intercept	17.02*	4.94 (0.50)*	0.92 (0.02)*	−0.04 (0.01)*	0.59 (0.05)*	−0.06 (0.04)	0.65 (0.10)*	0.04 (0.09)	0.27 (0.10)*	−0.08 (0.09)	0.08 (0.02)*	0.04 (0.01)*
	Slope	0.93	−0.02 (0.06)	0.82 (0.09)*	0.03 (0.04)	0.63 (0.19)*	0.06 (0.11)	0.39 (0.39)	−0.06 (0.23)	0.43 (0.37)	0.09 (0.22)	0.18 (0.09)*	−0.03 (0.04)
PIQ	Intercept	14.22*	3.63 (0.48)*	0.91 (0.02)*	−0.05 (0.01)*	0.58 (0.05)*	−0.07 (0.04)	0.67 (0.11)*	0.04 (0.09)	0.24 (0.11)*	−0.08 (0.09)	0.09 (0.02)*	0.05 (0.01)*
	Slope	1.13	−0.12 (0.07)	0.90 (0.26)*	0.12 (0.12)	1.06 (0.38)*	0.01 (0.18)	−0.32 (0.78)	0.23 (0.42)	1.23 (0.72)	−0.11 (0.37)	0.10 (0.26)	−0.12 (0.12)
VIQ	Intercept	5.15	5.33 (0.52)*	0.91 (0.02)*	−0.03 (0.01)*	0.61 (0.05)*	−0.03 (0.04)	0.59 (0.10)*	0.00 (0.08)	0.32 (0.10)*	−0.03 (0.07)	0.10 (0.02)*	0.03 (0.01)*
	Slope	0.71	0.00 (0.08)	0.81 (0.10)*	0.03 (0.04)	0.61 (0.17)*	0.04 (0.08)	0.40 (0.36)	−0.02 (0.17)	0.41 (0.33)	0.04 (0.16)	0.19 (0.10)*	−0.03 (0.04)

Wald: result of Wald chi-square test with two degrees of freedom. ME: main effect. Parameter estimates presented as value (standard error).

* $p < .05$.

first longitudinal G x SES interaction study to utilize more than two time points, and the first study to examine G x SES interaction longitudinally beyond early childhood. Pooling IQ measurements from up to five time points between ages 7 and 15 years, we observed significant SES moderation of the heritability of mean-level IQ and PIQ (intercept). These findings are consistent with our cross-sectional multivariate results, and with the results of previous cross-sectional studies that found significant interaction effects in middle childhood (Turkheimer et al., 2003) and adolescence (Harden et al., 2007; Rowe et al., 1999). Also consistent with Turkheimer et al. (2003), we did not observe a significant effect on VIQ intercept. However, significant interaction of verbal performance heritability and SES has been observed in another previous study of American twins (Rowe et al., 1999), and we found a significant VIQ effect at age 15 in univariate analyses. The extent to which SES modifies the heritability of some facets of intelligence more than others therefore remains unclear.

The fact that we did not observe a significant interaction of SES and the heritability of IQ slope could stem from our use of age-scaled scores, which are standardized to a mean of 100 and standard deviation of 15 at each age. In contrast to the variances of raw cognitive ability scores, which would be expected to increase across development, scaled score variances are by definition held constant over time. This invariance could have obscured slope interaction effects, should they exist, by limiting the extent to which children's scores could change between ages. Alternatively, it is possible that SES modifies the heritability of IQ starting point (intercept), but not the heritability of age-related changes in IQ (slope). This would diverge from the results of a study that observed significant interaction effects on change in mental ability in infancy (Tucker-Drob et al., 2011), and perhaps indicate that G x SES interaction effects on slope are present only in the early stages of cognitive development. Future studies will be needed to resolve this definitively.

Using age-scaled scores in longitudinal models of cognitive ability also introduces concerns about possible confounding via the Flynn effect, which holds that later generations exhibit systematically higher cognitive ability performance compared to earlier ones (Flynn, 1984). In the particular case of the LTS, wherein children were tested longitudinally for over a decade, one risk is that over time, the Flynn effect created differences in cognitive ability between our study sample and the historical WISC standardization sample such that the former's raw scores were systematically higher than the latter. Age-scaling scores in such a case could introduce cohort effects. This possibility is partially mitigated by the use of three versions of the WISC over the course of the LTS; age-scaled scores in each subsequent WISC version were normed on contemporary samples, and new WISC versions were adopted by the LTS shortly after their publication. Nevertheless, the use of scaled scores in longitudinal analyses remains a limitation.

Several other limitations should be considered when interpreting our results. First, our sample was of average SES, lacking substantial numbers of children raised in poverty. Given evidence that G x SES interaction may not be present in samples that have more universal access to enriching environmental resources (Tucker-Drob & Bates, 2016), analyzing a lower SES sample may have increased our likelihood of observing significant interaction effects. Second, the SES measure we used is a broad composite measure of environmental quality. The results of this study therefore do not clarify which specific environmental factors drive G x E interaction on cognitive ability, although others have investigated that exact question (for a review, see Hackman, Farah, & Meaney, 2010). Finally, even with the power boost offered by a multivariate approach, our power was less than ideal due to limited sample size, and we were unable to perform cross-sectional analyses at age 12.

This study adds to the substantial body of literature on G x SES interaction on cognitive ability among U.S. samples. It is now clear that merely partitioning the variance of cognitive performance into ACE components does not appreciate the complex interplay of genetic and environmental factors driving cognitive development. Existing G x E

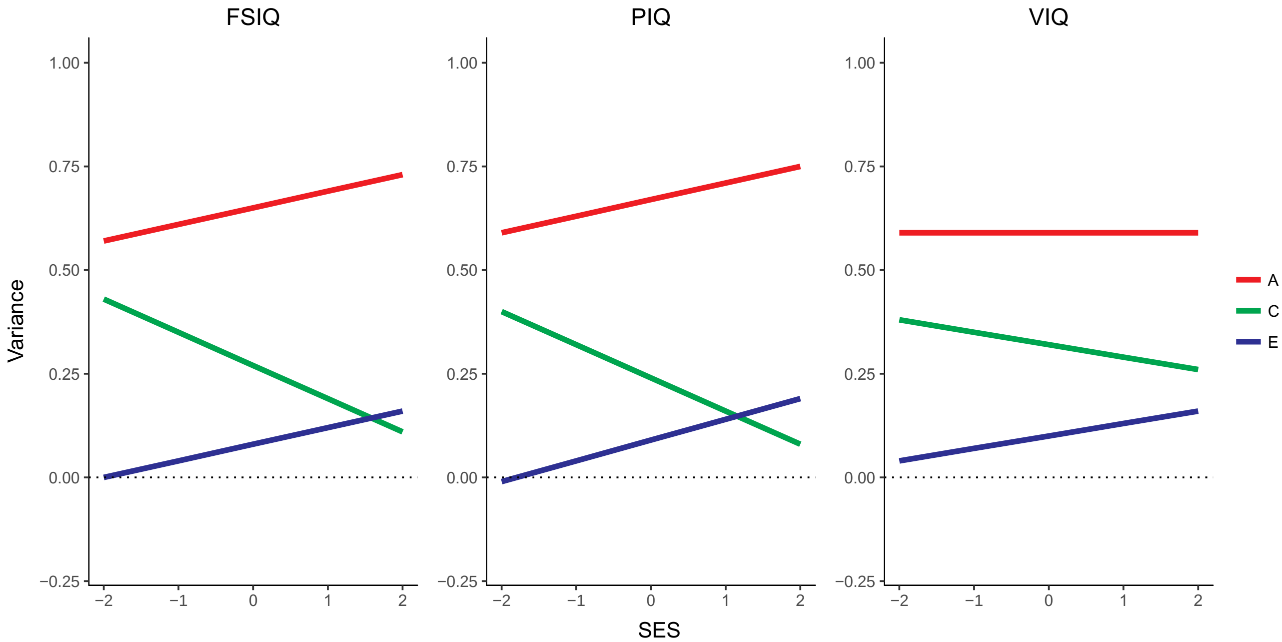

Fig. 7. Latent growth curve intercept ACE results. Variance: proportion of variance in cognitive ability attributable to A, C, and E (presented in red, green, and blue, respectively). SES is standardized to a mean of 0 and standard deviation of 1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

interaction studies, however, have not been as statistically robust as desired (Tucker-Drob & Bates, 2016). Utilizing multivariate methods as we have done here may help address this limitation and, in the case of longitudinal analyses, provide valuable insight into how G x SES interaction unfolds over the life course.

Funding

This work was supported by the National Institute on Aging (R03AG048850-01) and the John Templeton Foundation's Genetics and Human Agency initiative.

Declaration of competing interest

None.

Appendix A. Supplementary tables

Supplementary tables to this article can be found online at https://doi.org/10.1016/j.intell.2019.101400.

References

Allison, D. B., Allison, R. L., Faith, M. S., Paultre, F., & Pi-Sunyer, F. X. (1997). Power and money: Designing statistically powerful studies while minimizing financial costs. *Psychological Methods, 2*(1), 20–33. https://doi.org/10.1037/1082-989X.2.1.20.
Bates, T. C., Lewis, G. J., & Weiss, A. (2013). Childhood socioeconomic status amplifies genetic effects on adult intelligence. *Psychological Science, 24*(10), 2111–2116. https://doi.org/10.1177/0956797613488394.
Beam, C. R., & Turkheimer, E. (2013). Phenotype–environment correlations in longitudinal twin models. *Development and Psychopathology, 25*(1), 7–16. https://doi.org/10.1017/S0954579412000867.
Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology, 53*(1), 371–399. https://doi.org/10.1146/annurev.psych.53.100901.135233.
Figlio, D. N., Freese, J., Karbownik, K., & Roth, J. (2017). Socioeconomic status and genetic influences on cognitive development. *Proceedings of the National Academy of Sciences of the United States of America, 114*(51), 13441–13446. https://doi.org/10.1073/pnas.1708491114.
Fischbein, S. (1980). IQ and social class. *Intelligence, 4*(1), 51–63. https://doi.org/10.1016/0160-2896(80)90006-9.
Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95*(1), 29–51. https://doi.org/10.1037/0033-2909.95.1.29.
Grant, M. D., Kremen, W. S., Jacobson, K. C., Franz, C., Xian, H., Eisen, S. A., & Lyons, M. J. (2010). Does parental education have a moderating effect on the genetic and environmental influences of general cognitive ability in early adulthood? *Behavior Genetics, 40*(4), 438–446. https://doi.org/10.1007/s10519-010-9351-3.
Grasby, K. L., Coventry, W. L., Byrne, B., & Olson, R. K. (2017). Little evidence that socioeconomic status modifies heritability of literacy and numeracy in Australia. *Child Development.*. https://doi.org/10.1111/cdev.12920.
Hackman, D. A., Farah, M. J., & Meaney, M. J. (2010). Socioeconomic status and the brain: Mechanistic insights from human and animal research. *Nature Reviews Neuroscience, 11*, 651–659. https://doi.org/10.1038/nrn2897.
Harden, K. P., Turkheimer, E., & Loehlin, J. C. (2007). Genotype by environment interaction in adolescents' cognitive aptitude. *Behavior Genetics, 37*(2), 273–283. https://doi.org/10.1007/s10519-006-9113-4.
Hollingshead, A. B. (1975). *Four factor index of social status.* Unpublished manuscriptNew Haven, CT: Department of Sociology, Yale University.
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. https://doi.org/10.1007/s10519-004-3876-2.
Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Authorhttps://doi.org/10.13155/29825.
Purcell, S. (2002). Variance components models for gene-environment interaction in twin analysis. *Twin Research, 5*(6), 554–571. https://doi.org/10.1375/136905202762342026.
R Core Team (2018). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.r-project.org.
Rhea, S. A. (2015). Reviving the Louisville Twin Study: An introduction. *Behavior Genetics, 45*(6), 597–599. https://doi.org/10.1007/s10519-015-9763-1.
Rhemtulla, M., & Tucker-Drob, E. M. (2012). Gene-by-socioeconomic status interaction on school readiness. *Behavior Genetics, 42*(4), 549–558. https://doi.org/10.1007/s10519-012-9527-0.
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*(5), 1151–1162. https://doi.org/10.1111/1467-8624.00084.
Scarr-Salapatek, S. (1971). Race, social class, and IQ. *Science, 174*(4016), 1285–1295. https://doi.org/10.1126/science.174.4016.1285.
Schmitz, S., Cherny, S. S., & Fulker, D. W. (1998). Increase in power through multivariate analyses. *Behavior Genetics, 28*(5), 357–363. https://doi.org/10.1023/A:1021669602220.
Tucker-Drob, E. M., & Bates, T. C. (2016). Large cross-national differences in gene × socioeconomic status interaction on intelligence. *Psychological Science, 27*(2), 138–149. https://doi.org/10.1177/0956797615612727.
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. https://doi.org/10.1177/0956797610392926.
Turkheimer, E., Beam, C. R., & Davis, D. W. (2015). The Scarr-Rowe interaction in complete seven-year WISC data from the Louisville Twin Study: Preliminary report.

Behavior Genetics, 45(6), 635–639. https://doi.org/10.1007/s10519-015-9760-4.
Turkheimer, E., Beam, C. R., Sundet, J. M., & Tambs, K. (2017). Interaction between parental education and twin correlations for cognitive ability in a Norwegian conscript sample. *Behavior Genetics, 47*(5), 507–515. https://doi.org/10.1007/s10519-017-9857-z.
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*(6), 623–628. https://doi.org/10.1046/j.0956-7976.2003.psci_1475.x.
Wechsler, D. (1949). *Wechsler intelligence scale for children.* New York, NY: The Psychological Corporation.
Wechsler, D. (1974). *Wechsler intelligence scale for children–Revised.* New York, NY: The Psychological Corporation.
Wechsler, D. (1991). *Wechsler intelligence scale for children–Third edition.* San Antonio, TX: The Psychological Corporation.
Wilson, R. S. (1983). The Louisville Twin Study: Developmental synchronies in behavior. *Child Development, 54*(2), 298–316.