Genetic and Environmental Stability of Intelligence in Childhood and Adolescence

Sanja Franić,[1] Conor V. Dolan,[1] Catherina E.M. van Beijsterveldt,[1] Hilleke E. Hulshoff Pol,[2] Meike Bartels,[1] and Dorret I. Boomsma[1]
[1]*Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands*
[2]*Neuroimaging Research Group, University Medical Center Utrecht, Utrecht, The Netherlands*

The present study examined the genetic and environmental contributions to the temporal stability of verbal, non-verbal and general intelligence across a developmental period spanning childhood and adolescence (5–18 years). Longitudinal twin data collected in four different studies on a total of 1,748 twins, comprising 4,641 measurement points in total, were analyzed using genetic adaptations of the simplex model. The heterogeneity in the type of instrument used to assess psychometric intelligence across the different subsamples and ages allowed us to address the auxiliary question of how to optimally utilize the existing longitudinal data in the context of gene-finding studies. The results were consistent across domains (verbal, non-verbal and general intelligence), and indicated that phenotypic stability was driven primarily by the high stability of additive genetic factors, that the stability of common environment was moderate, and that the unique environment contributed primarily to change. The cross-subscale stability was consistently low, indicating a small overlap between different domains of intelligence over time. The high stability of additive genetic factors justifies the use of a linear combination of scores across the different ages in the context of gene-finding studies.

■ **Keywords:** intelligence, longitudinal measures, genetic stability, environmental stability, simplex model, heritability, structural equation modeling

Intelligence is one of the most frequently studied human behavioral traits. Over the past century it has motivated research across a diverse range of fields, including not only the behavioral sciences, but also genetics, neuroscience, molecular biology, and economics. It is one of the strongest known determinants of major life outcomes, such as educational attainment, occupational success, health and longevity (Deary et al., 2004; Gottfredson, 1997; Gottfredson & Deary, 2004; Neisser et al., 1996; Schmidt & Hunter, 2004). Over the past several decades, developments in multivariate statistical modeling coupled with the availability of large data sets collected in twins and relatives have allowed for the examination of the genetic and environmental etiology of individual differences in intelligence, and the more recent advances in genotyping and DNA sequencing have enabled the search the for specific genetic variants underlying the observed variation (e.g., Benyamin et al., 2014; Davies et al., 2011; Franić et al., 2013; Najmabadi et al., 2011).

The findings emerging from twin and family studies have univocally indicated: (1) a role of genetic factors in the etiology of intelligence (e.g., Bouchard & McGue, 1981; Deary et al., 2006; Plomin & Spinath, 2004; Plomin et al., 2008), and (2) an age-dependent pattern of heritability, with individual differences in late adolescence and adulthood being more strongly influenced by genetic factors than those in childhood (the heritability estimates typically ranging from ∼20% in infancy to ∼40–50% in middle childhood and ∼60–80% in adulthood; e.g., Bartels et al., 2002; Bishop et al., 2003; Boomsma & van Baal, 1998; Deary et al., 2006; Haworth et al., 2009; Hoekstra et al., 2007; McGue et al., 1993; Petrill et al., 2004; Plomin, 1986; Polderman et al., 2006). Environmental factors that contribute to similarity between family members (e.g., shared family environment) typically decline in etiological relevance throughout childhood and adolescence, while environmental factors that

Sanja Franić, Conor V. Dolan, Catherina E.M. van Beijsterveldt, Hilleke E. Hulshoff Pol, Meike Bartels & Dorret I. Boomsma

facilitate differentiation between family members appear to play a persistently modest to moderate role (e.g., Bartels et al., 2002; Boomsma & van Baal, 1998; Haworth et al., 2009). The temporal stability of intelligence (i.e., the conservation of the rank order of individuals over time) is estimated to be fairly high, with around 45–60% of the variance in childhood being preserved over any given ∼2-year interval (e.g., Bartels et al., 2002). This continuity in the observed individual differences is attributable predominantly to genetic factors, that is, to the expression of a single set of genes throughout development (e.g., Bartels et al., 2002; Bishop et al., 2003; Eaves et al., 1986; Hoekstra et al., 2007; Petrill et al., 2004; Rietveld et al., 2003). In addition to contributing to stability, genetic factors also generate change: age-specific genetic factors emerge at different ages, partly accounting for the lack of complete temporal stability. Environmental influences shared among family members, insofar as they are relevant, contribute mostly to stability, whereas the unshared environment contributes predominantly to change.

The aim of the present study is to contribute to the existing body of literature by providing one of the most comprehensive examinations of the genetic and environmental etiology of the observed stability of intelligence to date. We analyzed longitudinal twin data collected in four different studies on a total of 1,748 twins, measured across a developmental period spanning childhood and adolescence (5–18 years of age). In contrast to many of the previous examinations of the genetic and environmental stability of intelligence (but see Hoekstra et al., 2007; Rietveld, et al., 2003), we examine the stability of verbal and non-verbal abilities separately. In addition, we examine the stability of general intelligence (e.g., Jensen, 1998; Spearman, 1904). Because the choice of the psychometric instrument used to assess intelligence is inevitably dependent on the age of the participant, and because we combined data from four different studies (comprising 14 different subprojects), there is considerable heterogeneity in the measurement instrument used to assess intelligence across the different samples and ages. This is not dissimilar to the situation in many other data registries, where longitudinal measures are often collected using different instruments across the life span. In twin registries in particular, this issue becomes especially prominent in the context of gene-finding studies (e.g., Flint, 2013; Goldstein et al., 2013; Visscher et al., 2012), where specific genetic variants contributing to the variation in the observed trait (i.e., the phenotype) are sought. Here, the definition of the 'observed trait', or phenotype, is of considerable relevance (e.g., van der Sluis et al. 2010): How does one define a single 'observed trait' to be used in the analyses, given multiple measures over time? The presence of longitudinal data collected using different psychometric instruments allows us to address the auxiliary question of how to optimally utilize the existing twin registry data on intelligence in the context of gene finding studies (i.e., to examine whether data summarization is likely to diminish

the power to detect genetic effects; see e.g. van der Sluis, Verhage, Posthuma, & Dolan, 2010; Minică, Dolan, Kampert, Boomsma, & Vink, 2014; Medland & Neale, 2010).

In summary, the present study aims to: (1) assess the observed stability of verbal abilities, non-verbal abilities, and general intelligence, and (2) study the observed stability as a function of the underlying genetic and environmental factors. The structure of the dataset allows for an evaluation of how the results replicate and integrate across the different samples, and the presence of measures collected using multiple psychometric instruments allows us to address the practical question of how to optimally utilize the existing data in the context of gene-finding studies. Although the terms 'intelligence' and 'cognitive ability' have each been given a multitude of definitions (e.g., Jensen, 1998; Spearman, 1904), in the present article we use the two terms interchangeably.

Materials and Methods

Sample

The data were obtained from the Young Netherlands Twin Register (YNTR; van Beijsterveldt et al., 2013). The YNTR is a population-based register of Dutch twins born after 1986, recruited at birth and measured longitudinally at ages 1 through 18. The sample consisted of 1748 twins (including 872 complete twin pairs; 399 monozygotic (MZ) and 473 dizygotic (DZ)), and comprised four longitudinally measured subsamples (sample sizes: 544, 226, 552, and 426 individuals). A detailed structure of the data is given in Figure S1 (see Supplementary Material). The twins were measured longitudinally at ages 5–18. This generated 4,641 data points in total: 1,946, 808, 1,076, and 811 data points were available for the four subsamples, respectively. 47.5% of the participants were male.

Measures

Cognitive abilities were assessed longitudinally, using the Revised Amsterdam Children Intelligence Test (RAKIT; Bleichrodt et al., 1984), Wechsler Intelligence Scale for Children (WISC-R and WISC-III; Sattler, 1992; Van Haasen et al., 1986; Wechsler et al., 2002), Raven's Standard and Advanced Progressive Matrices (SPM, APM; Raven, 1960; Raven et al., 1998), and the Wechsler Adult Intelligence Scale (WAIS; Stinissen et al., 1970; Wechsler, 1997), the choice of test being largely dependent on the age of the participants. Subscale scores were derived following the guidelines in the tests' manuals (Bleichrodt et al., 1984; Sattler, 1992; Stinissen et al., 1970; Van Haasen et al., 1986; Wechsler 1997; Wechsler et al., 2002): for RAKIT, a verbal (V) and a nonverbal (NV) score were defined; for the WISC and the WAIS, the Verbal Comprehension Index (VCI), Perceptual Organization Index (POI), and Freedom from Distractibility Index (FDI) were defined. For Raven's SPM and APM, the total score (defined as the total number of items answered correctly) was used in the analyses. Because the variances of

the subscale scores across the different tests were quite heterogeneous in magnitude, to ease subsequent computation we standardized by dividing each variable by the product of its standard deviation and $\sqrt{5}$. This resulted in variances of an equal order of magnitude across the different tests.

Analyses

Genetic covariance structure modeling (Martin & Eaves, 1977) is the application of structural equation modeling (Bollen, 1989; Kline, 2005) to data collected in genetically informative samples, such as samples of twins (Franić et al., 2012; Neale & Cardon, 1992). In the classical twin design, the sample consists of MZ and DZ twin pairs. DZ twins share 50% of their segregating genes on average, while MZ twins share nearly their entire genome (Falconer & Mackay, 1996; van Dongen et al., 2012). The covariance structure of the phenotypes (i.e., observed traits) is typically modeled as a function of latent factors representing several sources of individual differences: additive genetic (A), shared environmental (C), and individual-specific environmental (E) sources.[1] Additive genetic influences are modeled by one or more A factors, which represent the total additive effects of genes relevant to the phenotype. Based on quantitative genetic theory (Falconer & Mackay, 1996; Mather & Jinks, 1971), the A factors are expected to correlate 1 across MZ twins and 0.5 across DZ twins. Environmental influences affecting the phenotype of both twins in an identical way, thereby increasing their similarity beyond what is expected based on genetic resemblance alone, are represented by one or more C factors. Therefore, by definition, the C factors correlate unity across twins (regardless of zygosity). All environmental influences causing phenotypic differences among family members are represented by one or more E factors. Thus, by definition, the E factors are uncorrelated across twins. Assuming an ACE model, the expected covariance structure in a multivariate twin model is thus:

$$\begin{pmatrix} \sum_{11} & \sum_{12} \\ \sum_{21} & \sum_{22} \end{pmatrix} = \begin{pmatrix} \sum_A + \sum_C + \sum_E & r_A \sum_A + \sum_C \\ (r_A \sum_A + \sum_C)^t & \sum_A + \sum_C + \sum_E \end{pmatrix}$$

where, given p phenotypes, $\mathbf{\Sigma}_{11}$ ($\mathbf{\Sigma}_{22}$) is the $p \times p$ covariance matrix of twin 1 (twin 2), $\mathbf{\Sigma}_{12}$ ($\mathbf{\Sigma}_{21}$) is the twin 1 – twin 2 $p \times p$ covariance matrix, and $\mathbf{\Sigma}_A$, $\mathbf{\Sigma}_C$ and $\mathbf{\Sigma}_E$ are the additive genetic, shared environmental, and unique environmental $p \times p$ covariance matrices, respectively. The coefficient r_A is the correlation between the additive genetic factors in twin 1 and twin 2 (1 in MZ and 0.5 in DZ twins).

In the present study, the temporal stability of intelligence (i.e., the stability of individual differences in performance on intelligence tests over time) and the temporal stability of genetic and environmental influences on intelligence (i.e., the degree to which the observed stability is attributable to

the continuity of the genetic/environmental factors that affect intelligence over time) were modeled using the simplex model (Guttman, 1954; Jöreskog, 1970). An example of a simplex model is depicted in Figure 1. In this model, the data at occasion t ($t = 1 \ldots T$) are regressed on data at the preceding measurement occasion (t-1), and the regression coefficient $\beta_{t,t-1}$ obtained in this regression is used as an indicator of temporal stability. For instance, a high β in the regression of verbal abilities at age 7 on verbal abilities at age 5 would indicate that the individual differences in verbal abilities are highly stable across this age span, that is, that the rank order of individuals is largely preserved. Thus, the variance of a measure at a given time point is modeled as a function of factors that are stable over time (e.g., the variance at time point t is a function of the variance at time point t-1 and of the regression coefficient $\beta_{t,t-1}$: $\sigma^2_{t-1} * \beta_{t,t-1}^2$) and newly emerging factors that affect the phenotype at the given time point but were absent at the preceding time point. The variance of a measure at time point t can thus be expressed as: $\sigma^2_t = \beta_{t,t-1}^2 * \sigma^2_{t-1} + \zeta_t$, where ζ_t denotes the variance due to innovation. A high $\beta_{t,t-1}$ in combination with low ζ_t indicates high temporal stability; conversely, a low $\beta_{t,t-1}$ and a high ζ_t indicate low stability, implying that the factors relevant to the phenotype at time t-1 decrease in relevance by time t, and newly emerging factors gain in relevance.

In a simplex model with p observed variables, the expected $p \times p$ covariance matrix $\mathbf{\Sigma}$ equals $(\mathbf{I} - \mathbf{B})^{-1} \mathbf{\Psi} (\mathbf{I} - \mathbf{B})^{-1t}$, where I is a $p \times p$ identity matrix, $\mathbf{B}$ is a $p \times p$ matrix containing the autoregressive coefficients (βs) in the model, and $\mathbf{\Psi}$ is a $p \times p$ matrix containing the variances and covariances (for the first measurement occasion) and the residual variances and covariances (for all the subsequent measurement occasions) of the observed variables. Thus, for the first three measurement occasions in Figure 1:

$$\Psi = \begin{pmatrix} \sigma^2_{V5} & c_5 & 0 & 0 & 0 & 0 \\ c_5 & \sigma^2_{NV5} & 0 & 0 & 0 & 0 \\ 0 & 0 & \zeta_{V7} & c_7 & 0 & 0 \\ 0 & 0 & c_7 & \zeta_{NV7} & 0 & 0 \\ 0 & 0 & 0 & 0 & \zeta_{V10} & c_{10} \\ 0 & 0 & 0 & 0 & c_{10} & \zeta_{NV10} \end{pmatrix}$$

where σ^2 denotes variance, ζ denotes residual variance, and c denotes (residual) covariance. Further,

$$\mathbf{B} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_{V7V5} & \beta_{V7NV5} & 0 & 0 & 0 & 0 \\ \beta_{NV7V5} & \beta_{NV7NV5} & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_{V10V7} & \beta_{V10NV7} & 0 & 0 \\ 0 & 0 & \beta_{NV10V7} & \beta_{NV10NV7} & 0 & 0 \end{pmatrix}$$

where $\beta_{t,t-1}$ is the regression coefficient in the linear regression of a variable at time t on a variable at time t-1 (e.g., β_{V7V5} denotes the regression of variable V_7 on variable V_5).

---

[1] In addition, the trait may be influenced by non-additive genetic factors (D), which include genetic interactions within the same locus (genetic dominance) or across different genetic loci (epistasis). In the present article, non-additive genetic effects were not modeled because the classical twin design does not allow for the simultaneous estimation of A, C, and D effects, and both the existing literature and our preliminary analyses favored an ACE over an ADE model.

Sanja Franić, Conor V. Dolan, Catherina E.M. van Beijsterveldt, Hilleke E. Hulshoff·Pol, Meike Bartels & Dorret I. Boomsma
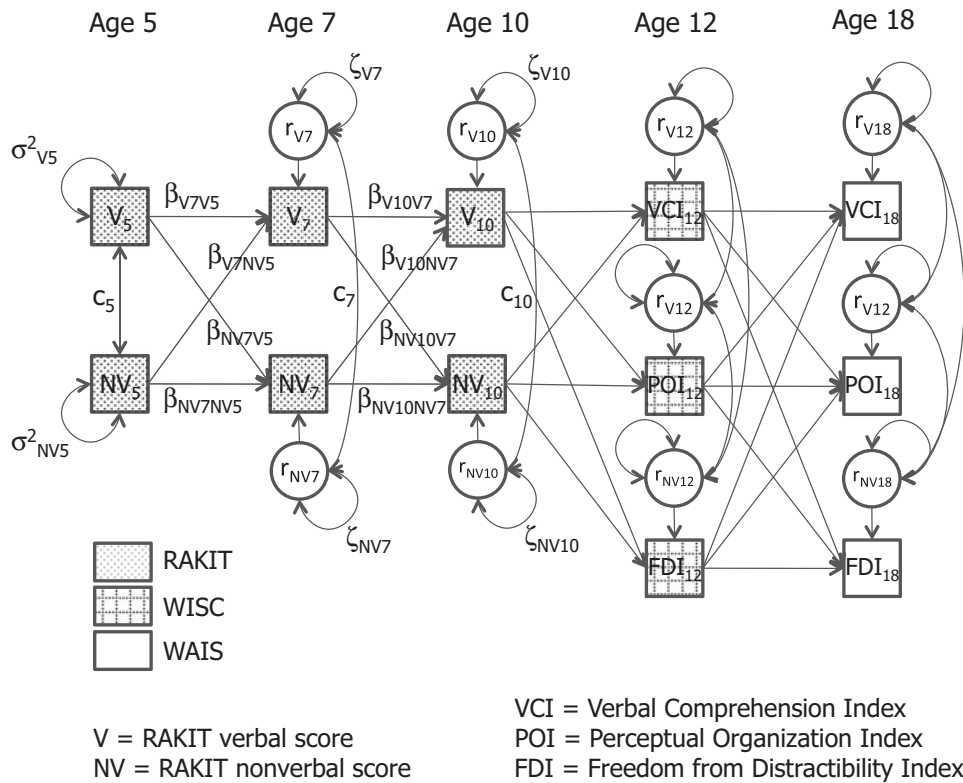
FIGURE 1

Phenotypic simplex model fitted to the data in Sample 1. Note: Subscale scores on the RAKIT, WISC, and WAIS at five measurement occasions are modeled. For simplicity, parameter notation is only given for the first three measurement occasions. σ^2 = variance, ζ = residual variance, c = (residual) covariance, β = regression coefficient. 'c' denotes covariance (between V_5 and NV_5) at the first measurement occasion, and residual covariance (i.e., covariance between the innovation factors) at subsequent measurement occasions.

To assess the contributions of genes and the environment to the observed stability and change in intelligence scores, a genetic adaptation of the simplex model was used (Boomsma & Molenaar, 1987; Boomsma et al., 1989; Franić et al., 2012; Neale & Cardon, 1992). In genetic adaptations of the simplex model, in contrast to modeling a single time series, the phenotype is modeled as a function of several (genetic and environmental) latent time series. For instance, in a model containing only additive genetic and unique environmental latent factors (AE model; Figure 2[2]), the phenotypic variable V measured at age t, V_t, is related to the additive genetic and unshared environmental factors A_t and E_t (t = 1,...,T), and simplex models, or first order autoregressions, are specified to account for the stability and change at the level of A and E (e.g., $\sigma^2_{Et} = \beta_{Et-1}{}^2 * \sigma^2_{Et-1} + \zeta_{Et}$).[3] The expected covariance structure of

the phenotype(s) is thus:

$$\begin{pmatrix} \sum_{11} & \sum_{12} \\ \sum_{21} & \sum_{22} \end{pmatrix} = \begin{pmatrix} \sum_A + \sum_E & r_A \sum_A \\ r_A \sum_A & \sum_A + \sum_E \end{pmatrix}$$

where (assuming the latent factors are expressed on the same scale as the phenotype) the covariance matrices $\mathbf{\Sigma}_A$ and $\mathbf{\Sigma}_E$ are modeled as follows:

$$\sum{}_A = (\mathbf{I} - \mathbf{B}_A)^{-1} \Psi_A (\mathbf{I} - \mathbf{B}_A)^{-1t},$$
$$\sum{}_E = (\mathbf{I} - \mathbf{B}_E)^{-1} \Psi_E (\mathbf{I} - \mathbf{B}_E)^{-1t},$$

This means that one can assess the contributions of genetic and environmental factors to the observed stability and the change in stability. The phenotypic covariance between consecutive time points may be due to genetic influences ($\beta_{At-1} \neq 0$), environmental influences ($\beta_{Et-1} \neq 0$), or both ($\beta_{At-1} \neq 0$ and $\beta_{Et-1} \neq 0$). Likewise, any lack of stability may be due to either or both sources of individual differences. For instance, intermediate phenotypic stability (e.g., a correlation of 0.5) may be due to perfect

---

[2] The figure assumes that the variances of the latent innovation factors (ζ_A and ζ_E) are scaled to 1.

[3] Here, σ^2_{Et} and σ^2_{Et-1} are the variances of the unique environmental factors at times t and t-1, β_{Et-1} is the regression of the E factor at time t on the E factor at t-1, and ζ_{Et} is the variance of the unique-environmental residual (i.e., innovation) at time t.
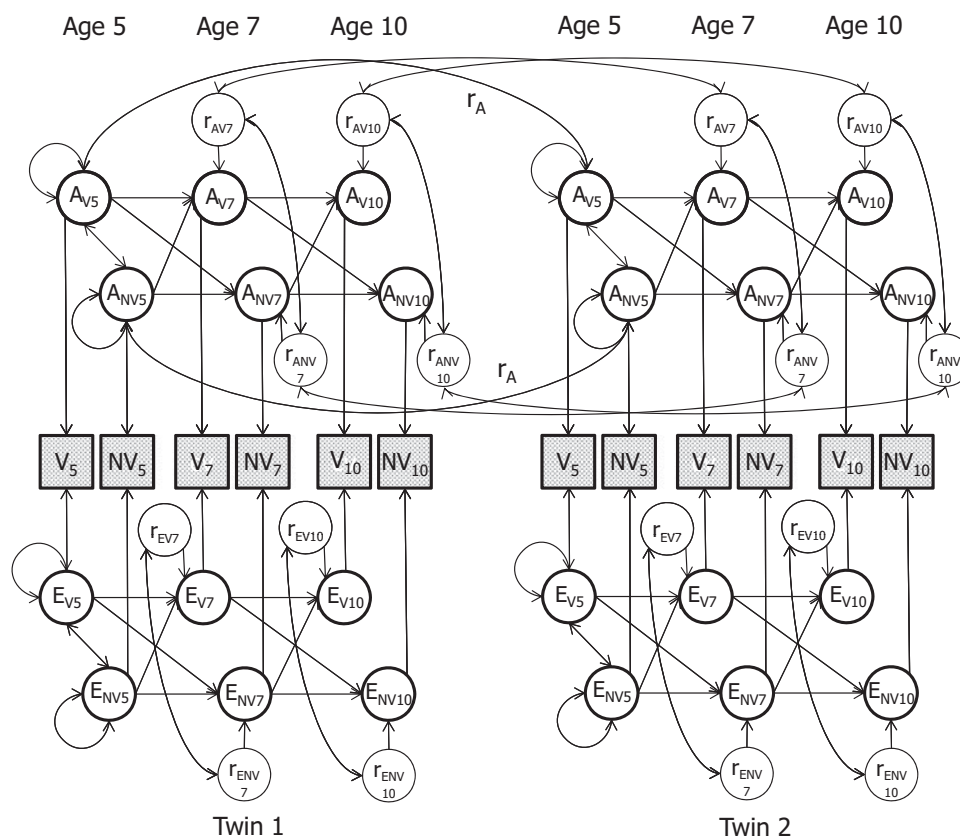
FIGURE 2

An example of an AE simplex model. Note: Data observed at three measurement occasions are modeled as a function of additive genetic and unique environmental factors (A and E, respectively), and simplex models are specified to account for the stability and change at the level of A and E.

genetic stability ($\zeta_{At} = 0$), in combination with complete environmental instability ($\beta_{Et,t-1} = 0$).

The present analyses were designed to examine the degree of phenotypic stability of intelligence, and assess the contributions of genes and the environment to the observed stability and change. This was achieved by fitting simplex models (described in the Supplementary Material) to intelligence tests subscale scores: RAKIT V and NV, WISC and WAIS VCI, POI, and FDI, and Raven sum scores. In addition to modeling the subscale scores, the stability of general cognitive ability (i.e., g; Jensen, 1998; Spearman, 1904) was assessed. The g factor was defined as a first-order factor underlying performance on the different subscales at a given age, and the temporal stability of g was examined on both the phenotypic, and the genetic and environmental level. Thus, overall, four different types of models were fitted: (a) phenotypic simplex models, (b) phenotypic simplex models with a g factor, (c) ACE simplex models, and (d) ACE simplex models with a g factor. These models were fitted to each of the four samples separately, resulting in 16 distinct sets of results. To accommodate for any possible mean differences across the sexes,

means were modeled separately for males and females in all analyses.

Results

For concision, the results pertaining to Sample 1 are presented in detail, while the results pertaining to the other three samples are summarized and discussed in view of their compatibility with the results in Sample 1. The full list of results (i.e., the parameter estimates obtained for all the four samples) is given in the Supplementary Material. Figure 3 displays the results obtained for Sample 1. For ease of interpretation, the results we present are fully standardized; that is, the variance of each (observed and latent) variable is 1. Stability is expressed as the proportion of variance of a variable at age t explained by the variables at t-1; this proportion is easily obtainable by subtracting the magnitude of innovation variance from the total variance, that is, as 1-ζ_t. A different standardization (allowing for a comparison of the relative magnitude of the A, C, and E variance components) is presented at the end of the Results section. The stability of the different subscales at a given age was largely
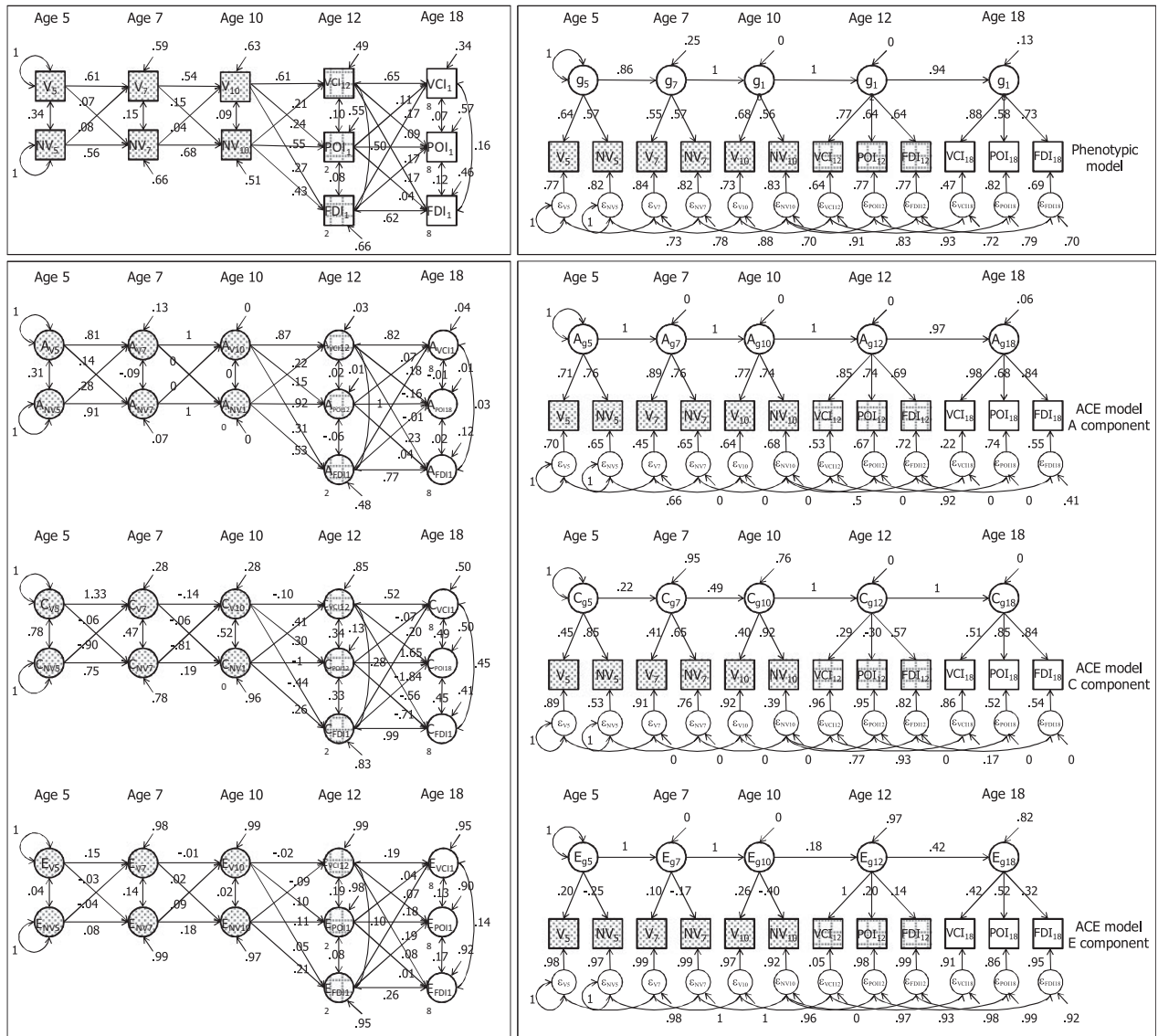
Sanja Franić, Conor V. Dolan, Catherina E.M. van Beijsterveldt, Hilleke E. Hulshoff Pol, Meike Bartels & Dorret I. Boomsma

FIGURE 3

Parameter estimates obtained for Sample 1. Note: Top left: phenotypic simplex model; top right: phenotypic simplex model with a *g* factor; bottom left: ACE simplex model; bottom right: ACE simplex model with a *g* factor. The results are completely standardized, that is, the total variance of each (latent and observed) variable in the model is 1. In the right panels, the numbers in the bottom of the figures denote residual innovation variance (ζ), rather than residual regression coefficients (βs). The residual βs are not depicted, but may be inferred from the residual variances (i.e., ζs). To minimize clutter in the figure, residual covariances are depicted as double-headed arrows connecting the observed variables (or the genetic/environmental components thereof), rather than the residuals.

comparable; thus, whenever possible, we describe general trends. When this is not warranted, we address the stability of the subscales separately.

Phenotypic Simplex Model

The temporal stability of intelligence subscales, as assessed using a phenotypic simplex model (upper left panel, Figure 3), is in the intermediate range, varying from 34% to 66% in Sample 1. Averaging over the subscale stabilities at each given age gives the mean stabilities of 38%, 43%, 43%, and 54% at the age intervals 5–7, 7–10, 10–12, and

12–18, respectively, indicating that the phenotypic stability of intelligence increases with age. This is especially evident if one considers that the time interval between the last two measurement points (ages 12 to 18) is more than twice the average time interval between the remaining consecutive measurement points, and that the correlation between measurement points is expected to decrease as an exponential function of their temporal distance, given equal stability over time. Thus, with stability being constant over age, one would expect a drop in the stability estimate from the observed 43% in the 10–12 interval to around 3.5% in the

12–18 interval; however, the actual stability estimate in the 12–18 interval is a high 54%, indicating a sharp increase in stability over this period. The cross-lag regression coefficients (e.g., RAKIT V to RAKIT NV) were generally small in magnitude compared to the main regression coefficients (e.g., RAKIT V to RAKIT V); estimates of variance explained by any single cross-lagged relationship ranged from 0.2% to 5.8% (see Figure 3 for estimates). Notably, the stability remained moderate to high despite the use of different tests (RAKIT, WISC, and WAIS).

In Sample 2, the average subscale stability at age 12 was 40%; an estimate comparable to the 43% stability at the same age interval in Sample 1. In Sample 3, the average subscale stabilities between the ages of 5–12 and 12–17 were 18% and 44%, respectively. An estimate of 18% in the 7-year interval prior to age 12 implies that, were the time intervals equal to those in Sample 1 (an average of 2.3 years prior to age 12), the stability would be estimated at 42%; highly consistent with the estimate obtained in Sample 1. The estimate of 44% in the 5-year interval between the ages of 12 and 17 implies that the stability would equal 58%, given a test-retest interval comparable to that of Sample 1 (i.e., 2.5 years). Thus, the temporal stability of intelligence as estimated in Sample 3 increases with age, and is consistent in both its magnitude and its observed increase with that estimated in Samples 1 and 2. In Sample 4, the mean subscale stability between the ages 15 and 18 is estimated at 30%. This is lower than the estimates obtained for the other samples; however, in Sample 4 the Raven sum score alone is used as a predictor of the three WAIS subscales (Figure S3 in Supplementary Material). Thus, while the 30% estimate may reflect a lower temporal stability, it may also be attributable to the relatively low correlation between the WISC subscales and the Raven.

Overall, the phenotypic subscale analyses indicate moderate to high stability of individual differences in intelligence across childhood and adolescence. The stability increases with age; i.e., the individual differences in intelligence become increasingly stable as individuals transition from childhood to adolescence. Notably, the stability remains in the intermediate to high range despite the variation in the instruments used to assess intelligence, and the results replicate well despite the differences in tests and measurement intervals across the four different samples.

Phenotypic Simplex Model With a g Factor

The upper right panel of Figure 3 shows the phenotypic simplex model with a *g* factor fitted to Sample 1. On average, the *g* factor explained around 37%, 31%, 38%, 47%, and 55% of subscale variance at ages 5, 7, 10, 12, and 18, respectively (possibly indicating an increasing role of *g* in intelligence over time, but also possibly reflecting the differences in the tests used). The temporal stability of the *g* factor is remarkably high: nearly the entire inter-individual variation at a given age can be predicted by the variation at

the preceding age. The residual, subscale-specific variation displays a modest degree of stability over time: 20% on average. It should, however, be noted that this is a lower estimate of residual stability, as the estimates of the subscale-specific variation also include measurement error.

In Sample 2, the *g* factor explained an average of 47% and 37% of subscale variance at ages 9 and 12, respectively. The stability of *g* from age 9 to age 12 was around 80%, and the stability of the residual scores was modest (∼15%), as in Sample 1. In Sample 3, the stability estimates were somewhat lower (42% and 65% at intervals 5–12 and 12–17, respectively). Note, however, that stability estimates of 42% and 65% over 7 and 5 years, respectively, imply that the stability would have been estimated at around 75% and 80%, respectively, had the time intervals been comparable to those of Samples 1 and 2 (2–2.5 years). In Sample 4, the stability was estimated at 60%. Again, it should be borne in mind that in Sample 4, Raven alone was used as a predictor of all three WAIS subscales; therefore, the lower stability estimate may reflect a lower temporal stability in Sample 4, but may also be due to the relatively low correlation between the Raven sum score and the WAIS subscale scores.

In summary, the phenotypic stability of *g* over childhood and adolescence is high, and exceeds the stability of individual subscales. The *g* factor explains around 30–55% of subscale variance, regardless of the test used. Across all four samples, the stability of the subscale-specific scores is modest (around 15–20%).

ACE Simplex Model

The lower left panel of Figure 3 shows the parameter estimates obtained for the ACE simplex model fitted to Sample 1. As evident from the figure, the additive genetic influences on intelligence are highly stable; the stability estimates range from approximately 90% to 100%, and display an increase with age. Therefore, the genes that influence intelligence in early childhood overlap largely, if not entirely, with those that affect it throughout childhood and adolescence. Cross-lag regressions across measurement points and residual correlations within measurement points are fairly low, indicating that the genetic factors affecting verbal abilities are largely distinct from those affecting non-verbal abilities, both within and across measurement points. The stability of common environmental influences is in the intermediate range, and differs per subscale: the common environmental stability of verbal abilities before age 10 is considerably higher than the common environmental stability of non-verbal abilities in this period; at later ages, however, the difference in stability between the subscales appears to disappear. However, as the magnitude of C component is small and decreases over time (see end of the Results section), the apparent differences in subscale stabilities are likely attributable to the unreliability of the relevant parameters. The unique environmental

influences display virtually no stability over time; the stability estimates are close to zero at all time points.

In Sample 2, the additive genetic influences are highly stable (over 90% on average), the common environmental stability is high (~85% on average), and unique environmental stability is low (12% on average). Similarly, in Sample 3, the additive genetic stability is high (close to 100% except for the FOI subscale at age 12, the stability of which is estimated at 38%), the common environmental stability is estimated at 47% in the 5–12 interval and 100% in the 12–17 interval, and the E stability is virtually zero. In Sample 4, the A influences are estimated to be around 80% stable, the C influences around 60% stable, while the E influences display virtually zero stability.

Overall, the results indicate a high additive genetic stability (largely 90–100%), a moderate to high common environmental stability, and a complete absence of unique environmental stability for both verbal and non-verbal abilities. The cross-subscale (e.g., verbal-nonverbal) stability is consistently low.

ACE Simplex Model With a g Factor

The ACE simplex model with a g factor fitted to Sample 1 is shown in the lower right panel of Figure 3. In this sample, the additive genetic g factor explains around 60% of the additive genetic subscale variance and displays nearly perfect stability; 100% at most time points. Similarly, the additive genetic subscale residuals generally display a high temporal stability. The common environmental g factor explains around 40% of the common environmental subscale variance. The stability of common environmental influences appears to increase after age 10: the stability estimates are 5%, 24%, 100%, and 100% at ages 7, 10, 12, and 18, respectively. However, the magnitude of the C (and E) variance is relatively small, and thus the reliability of the C (and the E) stability parameters is likely low. Overall, the residual C stability is estimated to be high. The unique environmental component displays an opposite pattern to the common environmental component: the stability of g before the age 10 is high, and declines substantially thereafter. However, the unique environmental g factor explains only 16% of the unique environmental subscale variance on average; the rest is explained by the subscale-specific E factors, which display virtually no stability.

In Sample 2, the additive genetic g factor explains 52% of the additive genetic subscale variance, and is 70% stable on average. Similarly, the residuals are highly stable (85%). The C_g factor displays complete stability, and explains 90% of the C subscale variance. The E_g factor explains only 24% of the unique environmental subscale variance, and is 30% stable on average, with highly unstable residuals. In Sample 3, the additive genetic g factor explains 70% of the subscale variance and is 93–100% stable. The C_g factor explains around 60% of the C subscale variance, and declines in stability from 100% at ages 5–12 to 34% at ages 12–17. Again,

however, the variance in the C stability estimates is likely due to the small magnitude of C. The E subscale variance was only modestly explained by E_g (~25%), and displayed stability neither at the g level, nor at the residual level. In Sample 4, the A_g, C_g, and E_g factors explained around 76%, 52%, and 11% of their respective variance, and were 100%, 100%, and 16% stable, respectively.

In summary, the A_g, C_g, and E_g factors explained an average of ~65%, ~60%, and ~20% of the A, C, and E variance, respectively. The A_g factor was highly stable over time (mostly close to 100%), with highly stable residuals. The C_g factor was generally highly stable (close to 100%), with some exceptions (ages 5–10 in Sample 1 and ages 12–17 in Sample 3; however, considering the small magnitude of the C variance component, these exceptions likely reflect the unreliability of the estimates. The E_g factor displayed modest stability (around 35% on average), but explained only around 30% of the E variance, the remainder of the variance being entirely unstable (close to 0%) across all samples.

Magnitude of Variance Components

The relative magnitude of the A, C, and E variance components, as estimated in the ACE simplex models and averaged over subscales at each age, is depicted in Figure S2 (Supplementary Material). An age-related increase in heritability accompanied by a relative decline in common environmental variance, expected based on the literature (e.g., Bartels et al., 2002; Bishop et al., 2003; Boomsma & van Baal, 1998; Deary et al., 2006; Haworth et al., 2009; Hoekstra et al., 2007; McGue et al., 1993; Petrill et al., 2004; Plomin, 1986; Polderman et al., 2006), is evident in Samples 1, 3, and 4. In Sample 2, where only two measurement points were available (ages 9 and 12), this trend was not apparent. This lack of trend can presumably be attributed to the brevity of test-retest time interval.

Integrated Results

Figures 4 and 5 depict estimates of standardized variance components and A, C, and E stabilities, respectively, obtained across all four samples and shown for verbal and non-verbal abilities separately. Unlike data in Figure 5, the data in Figure 4 did not appear to show considerable deviations from linearity; therefore the general trends are represented by linear regression lines weighted by sample size in Figure 4 and by a smoothing function (lowess function as implemented in R; R Core Team, 2013) in Figure 5. Consistently with Figure S2, an increase in the relative magnitude of additive genetic variance accompanied by a decrease in common (and, to some extent, unique) environmental variance is evident from Figure 4. Figure 5 indicates an increase in stability of all three components over time, and suggests that the observed phenotypic stability is driven primarily by additive genetic factors, with unique environment contributing primarily to change. Note that, for comparability,
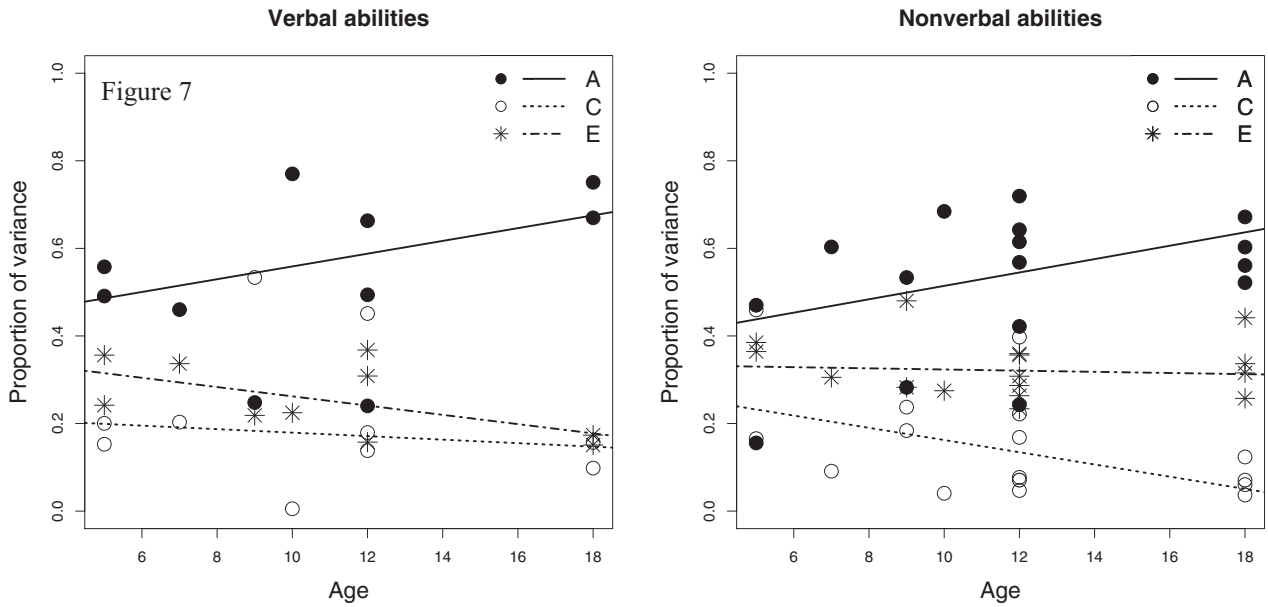
FIGURE 4

The relative magnitude of the A, C, and E variance components (y-axis) as a function of age, for verbal (left) and non-verbal (right) abilities. Note: All available estimates from the four samples are included. Regression lines (weighted by sample size) represent the general trends.
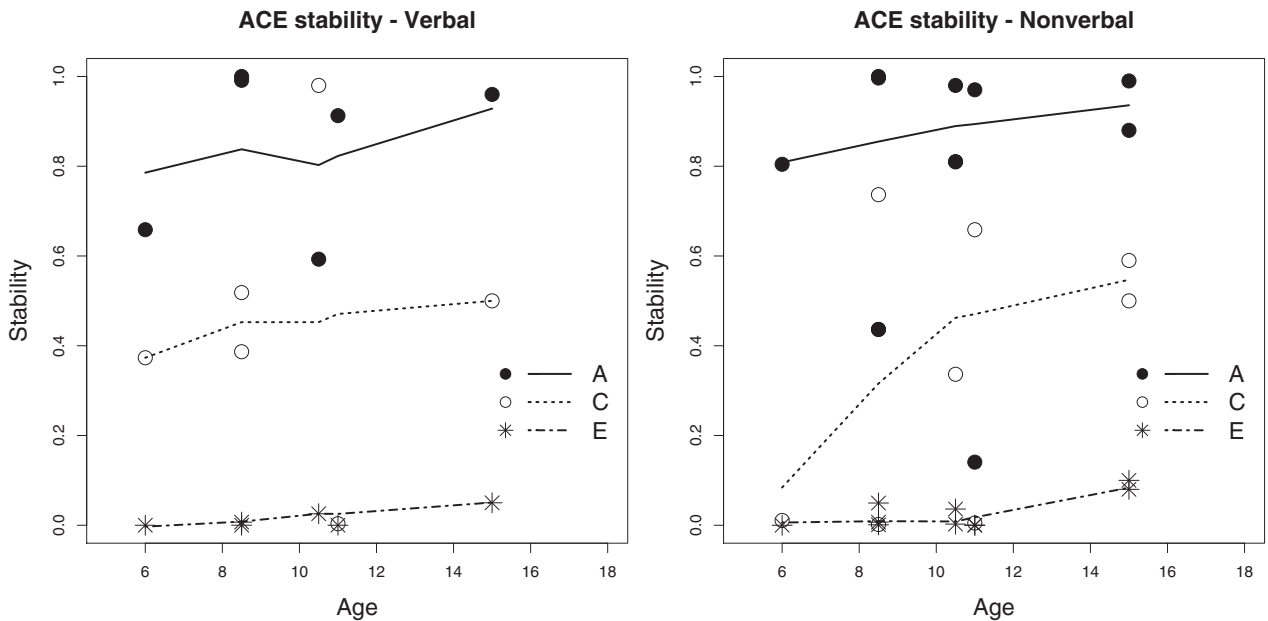

FIGURE 5

The ACE stability of verbal (left) and non-verbal (right) abilities. Note: All available estimates from the four samples are included, and re-expressed on a scale on which all measurement points are equidistant (6 years). Lines (locally weighted scatterplot smoothing functions) represent general trends.

Figure 5 re-expresses the stability estimates on a scale on which all measurement points are equidistant (6 years). As explained earlier (see Supplementary Material), stability estimates are dependent on the time interval one uses for estimation and therefore the absolute magnitude of the stability estimates is not interpretable in itself. The choice of time interval used to re-express the estimates is therefore arbitrary; the reason a 6-year period was chosen in this case is the fact that, with smaller (e.g., 1-year) time intervals, the stability estimates reach an upper bound, making it impossible to

Sanja Franić, Conor V. Dolan, Catherina E.M. van Beijsterveldt, Hilleke E. Hulshoff Pol, Meike Bartels & Dorret I. Boomsma

TABLE 1

The Phenotypic, Genetic and Environmental Correlations Obtained in the Four Samples Under an ACE Simplex Model

Σ_A	5	7	9	10	12	15	17	18	Σ_C	5	7	9	10	12	15	17	18
5		0.96		0.95	0.79			0.82	5		0.28		0.04	0			0
7	0.89			1	0.84			0.81	7	0.63			0.18	-0.01			0
9					0.94				9					0.59			
10	0.90	1			0.83			0.84	10	-0.07	-0.29			-0.40			0.36
12	0.91	0.97	0.75	0.97				0.95	12	-0.29	0.02	.95	0.06				0.20
15									15								
17									17								
18	0.83	0.94		0.93	0.96				18	00	0.04		-0.03	0.70			
Σ_E	5	7	9	10	12	15	17	18	Σ	5	7	9	10	12	15	17	18
5		0.08		0.01	0.04			0	5		0.58		0.42	0.33			0.20
7	0.15			0.18	0.04			0.01	7	0.64			0.70	0.43			0.33
9					0.09				9					0.58			
10	0	0			0.17			0.05	10	0.38	0.59			0.58			0.44
12	0.04	0	0.14	-0.03				0.21	12	0.34	0.42	.69	0.69				0.68
15									15								
17									17								
18	0	0		0	0.21				18	0.23	0.36		0.57	0.79			

Note: Note: The correlations are given for verbal (below diagonal) and non-verbal (above diagonal) abilities separately.

distinguish between the stability of the different variance components (i.e., the C and E stability estimates increase, whereas the A stability estimates readily hit the upper bound of 1). Finally, Table 1 gives all available estimates of the phenotypic, genetic and environmental correlations obtained under an ACE simplex model, for verbal and non-verbal abilities separately. Again, it is evident that the observed stability of intelligence is driven primarily by additive genetic factors, with common environment contributing both to stability and change, and the unique environment predominantly generating change.

Discussion

The present study examined the stability of verbal abilities, non-verbal abilities, and general intelligence across childhood and adolescence, and assessed the genetic and environmental etiology of this stability. Other questions included the feasibility of combining results on multiple types of intelligence tests administered in a longitudinal design with the aim of utilizing the combined score in the context of gene-finding studies, and the relationship between different types of intellectual abilities over time (and the genetic/environmental etiology thereof).

The results indicate an intermediate to high phenotypic stability of individual differences in intelligence across the developmental period under study, with an increase in stability as individuals transition from childhood to adolescence. General intelligence, defined as a first first-order latent factor underlying subscale performance at a given age, explained around 30–55% of variance in subscale performance and displayed high temporal stability, exceeding that of individual subscales. The phenotypic stability appears to be driven primarily by genetic factors: the additive genetic influences were highly to entirely stable. The environment shared by family members appeared to contribute to sta-

bility to a moderate degree, while environmental factors unique to family members contributed mainly to innovation (i.e., to temporal instability). Similarly, the observed stability in the *g* factor was driven primarily by genetic factors: the additive genetic *g* factor displayed near complete stability, the common environmental *g* factor was generally stable but explained less of the phenotypic variance than the A_g factor, while the unique environmental *g* factor was modestly stable but explained only a minor fraction of the phenotypic variance in *g*. An age-related increase in heritability accompanied by a relative decline in common environmental variance, expected based on the literature (e.g., Bartels et al., 2002; Bishop et al., 2003; Boomsma & van Baal, 1998; Deary et al., 2006; Haworth et al., 2009; Hoekstra et al., 2007; McGue et al., 1993; Petrill et al., 2004; Plomin, 1986; Polderman et al., 2006), was observed. In addition, the cross-subscale stability was consistently low, indicating a small to non-existent contribution of one domain of intelligence to another over time.

The stability of intelligence remained in the intermediate to high range despite the variation in the instruments used to assess it, and the results replicated well across the four samples despite the variation in tests and the time intervals used for estimation. The former relates to a common situation in data registries (e.g., twin registries), where data are often collected using a number of different psychometric instruments, the choice of test often being dependent on participants' age. Given the increased accessibility of genotyping and sequencing technologies and the consequent increase in the use of twin registry data in gene-finding studies, the question of how to optimally combine the existing longitudinal data in defining the phenotype for such studies is gaining in relevance. In this context, there are two prominent issues: (1) the actual modeling of a measured genetic variant in multivariate data; and (2) the accommodation of family members in the analysis. The latter does

not pose a problem as the methods and software for family-based gene finding studies are well developed (e.g., Chen & Abecasis, 2007; Lippert et al., 2011; Minică et al., 2013, 2014; Purcell et al., 2007). The former is potentially more problematic as full multivariate phenotypic modeling of family data is not computationally feasible, or perhaps even desirable. There are many possible loci of a genetic variant effect in a multivariate model, and therefore many possible models to consider. The present results, as pertaining to the longitudinal genetic covariance structure, suggest that a simple phenotypic sum score based on the repeated measures within a cognitive domain (e.g., verbal) should not result in any appreciable loss of information in a genetic association study (see Minică et al., 2010). Whether one should sum over cognitive domains is a different question. The genetic *g* factor accounted for about 60% of the genetic variance of the subtest scores. Summing over domains will only improve the power to detect a genetic variant if it contributes to this common genetic variance. Rather than running the risk of missing genetic variants, it is advisable to carry out gene-finding studies for each domain separately. One can still arrive at an omnibus test of the genetic variant (i.e., address the question of whether the genetic effect generalizes over domains) by combining the statistical results (van der Sluis et al., 2013).

While we believe that the high genetic stability provides a reasonable justification for summing over repeated measures within an individual, we note that this recommendation is limited in two ways. First, it applies to the present longitudinal results as obtained in the repeated measures design. From the point of view of power, a cross-sectional design may be preferable (and is certainly more efficient and cheaper to implement). However, the exact relationship between power and design is beyond the present scope (Minică et al., 2010, do consider different multivariate designs). Second, the recommendation is based strictly on the present choice to model the covariance structure by means of autoregressive and cross-lagged modeling. This approach is informative with respect to stability, but does not consider developmental change from the point of view of individual growth curves (Ramsden et al., 2011). We did not consider growth curve modeling as our IQ test scores were age-corrected, meaning that the present data were not informative with respect to individual developmental growth curves. Finally, the results present here were based on the standard genetic simplex model, in which A, C and E are assumed to be uncorrelated sources of individual differences. Whether this assumption (e.g., the absence of genotype-environment covariance) is valid to a reasonable approximation is an open question. Any genotype-environment covariance is unlikely to undermine our recommendations concerning data summarization in gene-finding studies. However, a representation involving phenotype to environment transmission, typically envisaged as smart children contributing to their own 'smart' environment (a.k.a. niche

picking; Eaves et al., 1977) is possible (Dolan et al., 2014a, 2014b).

Supplementary Material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/thg.2014.26.

References

Bartels, M., Rietveld, M. J. H., Van Baal, G. C. M., & Boomsma, D. I. (2002). Genetic and environmental influences on the development of intelligence. *Behavior Genetics*, *32*, 237–249.

Benyamin, B., Pourcain, B., Davis, O., Davies, G., Hansell, N., Brion, M.-J., . . . Miller, M. (2014). Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. *Molecular Psychiatry*, *19*, 253–258.

Bishop, E., Cherny, S. S., Corley, R., Plomin, R., DeFries, J. C., & Hewitt, J. K. (2003). Development genetic analysis of general cognitive ability from 1 to 12 years in a sample of adoptees, biological siblings, and twins. *Intelligence*, *31*, 31–49.

Bleichrodt, N., Drenth, P., Zaal, J., & Resing, W. (1984). *Revisie Amsterdamse Kinder Intelligentie Test, RAKIT*. Lisse: Swets & Zeitlinger.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Boomsma, D. I., Martin, N. G., & Molenaar, P. C. M. (1989). Factor and simplex models for repeated measures: Application to two psychomotor measures of alcohol sensitivity in twins. *Behavior Genetics*, *19*, 79–96.

Boomsma, D. I., & Molenaar, P. C. M. (1987). The genetic-analysis of repeated measures. I. Simplex models. *Behavior Genetics*, *17*, 111–123.

Boomsma, D. I., & van Baal, G. C. M. (1998). Genetic influences on childhood IQ in 5- and 7-year-old Dutch twins. *Developmental Neuropsychology*, *14*, 115–126.

Bouchard, T. J., & McGue, M. (1981). Familial studies of intelligence: A review. *Science*, *212*, 1055–1059.

Chen, W.-M., & Abecasis, G. R. (2007). Family-based association tests for genomewide association scans. *The American Journal of Human Genetics*, *81*, 913–926.

Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S. E., Liewald, D., . . . Deary, I. J. (2011). Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry*, *16*, 996–1005.

Deary, I. J., Spinath, F. M., & Bates, T. C. (2006). Genetics of intelligence. *European Journal of Human Genetics*, *14*, 690–700.

Sanja Franić, Conor V. Dolan, Catherina E.M. van Beijsterveldt, Hilleke E. Hulshoff Pol, Meike Bartels & Dorret I. Boomsma

Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H. C. (2004). The impact of childhood intelligence on later life: following up the Scottish mental surveys of 1932 and 1947. *Journal of Personality and Social Psychology, 86*, 130–147.

Dolan, C. V., de Kort, J. M., Kan, K. J., van Beijsterveldt, C. E. M., Bartels, M., & Boomsma, D. I. (2014a). Can GE-covariance originating in phenotype to environment transmission account for the Flynn effect? *Journal of Intelligence.*

Dolan, C. V., de Kort, J. M., van Beijsterveldt, C. E. M., Bartels, M., & Boomsma, D. I. (2014b). GE Covariance through phenotype to environment transmission: An assessment in longitudinal twin data and application to childhood anxiety. *in press.*

Eaves, L. J., Last, K., Martin, N. G., & Jinks, J. L. (1977). A progressive approach to non-additivity and genotype-environmental covariance in the analysis of human differences. *British Journal of Mathematical and Statistical Psychology, 30*, 1–42.

Eaves, L. J., Long, J., & Heath, A. C. (1986). A theory of developmental change in quantitative phenotypes applied to cognitive development. *Behavior Genetics, 16*, 143–162.

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics* (4th ed.). Harlow, Essex, UK: Longmans Green.

Flint, J. (2013). GWAS. *Current Biology, 23*, R265–R266.

Franić, S., Dolan, C. V., Borsboom, D., & Boomsma, D. I. (2012). Structural equation modeling in genetics. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 617-635). New York: Guilford Press.

Franić, S., Groen-Blokhuis, M. M., Dolan, C. V., Kattenberg, M. V., Xiao, X., Scheet, P. A., . . . Boomsma, D. I. (2013). *IQ: Shared genetic basis between mendelian disorders and a polygenic trait.* Manuscript submitted for publication.

Goldstein, D. B., Allen, A., Keebler, J., Margulies, E. H., Petrou, S., Petrovski, S., & Sunyaev, S. (2013). Sequencing studies in human genetics: Design and interpretation. *Nature Reviews Genetics, 14*, 460–470.

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence, 24*, 79–132.

Gottfredson, L. S., & Deary, I. J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science, 13*, 1–4.

Guttman, L. (1954). A new approach to factor analysis: the Radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences.* Glencoe, IL: The Free Press.

Haworth, C., Wright, M., Luciano, M., Martin, N., De Geus, E., Van Beijsterveldt, C., . . . Davis, O. (2009). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular Psychiatry, 15*, 1112–1120.

Hoekstra, R. A., Bartels, M., & Boomsma, D. I. (2007). Longitudinal genetic study of verbal and nonverbal IQ from early childhood to young adulthood. *Learning and Individual Differences, 17*, 97–114.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger Publishers.

Jöreskog, K. G. (1970). Estimation and testing of simplex models. *British Journal of Mathematical and Statistical Psychology, 23*, 121–145.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press.

Lippert, C., Listgarten, J., Liu, Y., Kadie, C. M., Davidson, R. I., & Heckerman, D. (2011). FaST linear mixed models for genome-wide association studies. *Nature Methods, 8*, 833–835.

Martin, N. G., & Eaves, L. J. (1977). Genetic-analysis of covariance structure. *Heredity, 38*, 79–95.

Mather, K., & Jinks, J. L. (1971). *Biometrical genetics* (2nd ed.). London: Chapman and Hall.

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

Medland, S. E., & Neale, M. C. (2010). An integrated phenomic approach to multivariate allelic association. *European Journal of Human Genetics, 18*, 233–239.

Minică, C. C., Boomsma, D. I., Van Der Sluis, S., & Dolan, C. V. (2010). Genetic association in multivariate phenotypic data: Power in five models. *Twin Research and Human Genetics, 13*, 525–543.

Minică, C. C., Dolan, C. V., Hottenga, J.-J., Willemsen, G., Vink, J. M., & Boomsma, D. I. (2013). The use of imputed sibling genotypes in sibship-based association analysis: On modeling alternatives, power and model misspecification. *Behavior Genetics, 43*, 254–266.

Minică, C. C., Dolan, C. V., Kampert, M. M. D., Boomsma, D. I., & Vink, J. M. (2014). *Sandwich corrected standard errors in family-based genomewide association studies.* Manuscript submitted for publication.

Najmabadi, H., Hu, H., Garshasbi, M., Zemojtel, T., Abedini, S. S., Chen, W., . . . Ropers, H. H. (2011). Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature, 478*, 57–63.

Neale, M. C., & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families.* Dordrecht: Kluwer.

Neisser, U., Boodoo, G., Bouchard Jr, T. J., Boykin, A. W., Brody, N., Ceci, S. J., . . . Sternberg, R. J. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*, 77.

Petrill, S. A., Lipton, P. A., Hewitt, J. K., Plomin, R., Cherny, S. S., Corley, R., & DeFries, J. C. (2004). Genetic and environmental contributions to general cognitive ability through the first 16 years of life. *Developmental Psychology, 40*, 805–812.

Plomin, R. (1986). *Development, genetics, and psychology.* Hillsdale, NJ: Erlbaum.

Plomin, R., Defries, J. C., McClearn, G. E., & McGuffin, P. (2008). *Behavioral Genetics.* New York: W.H.Freeman & Co.

Plomin, R., & Spinath, F. M. (2004). Intelligence: Genetics, genes, and genomics. *Journal of Personality and Social Psychology, 86*, 112–129.

Polderman, T. J. C., Gosso, M. F., Posthuma, D., van Beijster-veldt, T. C., Heutink, P., Verhulst, F. C., & Boomsma, D. I. (2006). A longitudinal twin study on IQ, executive functioning, and attention problems during childhood and early adolescence. *Acta Neurologica Belgica*, *106*, 191–207.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Daly, M. J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*, 559–575.

Ramsden, S., Richardson, F. M., Josse, G., Thomas, M. S., Ellis, C., Shakeshaft, C., . . . Price, C. J. (2011). Verbal and non-verbal intelligence changes in the teenage brain. *Nature*, *479*, 113–116.

Raven, J. C. (1960). *Guide to the standard progressive matrices: Sets A, B, C, D and E*: HK Lewis. London: Lewis & Co.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 4, the advanced progressive matrices*. Oxford, England: Oxford Psychologists Press.

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Rietveld, M., Dolan, C. V., Van Baal, G., & Boomsma, D. I. (2003). A twin study of differentiation of cognitive abilities in childhood. *Behavior Genetics*, *33*, 367–381.

Sattler, J. M. (1992). *Assessment of children: WISC–III and WPPSI–R supplemental*. San Diego, CA: Author.

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: occupational attainment and job performance. *Journal of Personality and Social Psychology*, *86*, 162–173.

Spearman, C. (1904). 'General intelligence', objectively determined and measured. *The American Journal of Psychology*, *15*, 201–292.

Stinissen, J., Willems, P., Coetsier, P., & Hulsman, W. (1970). *Manual for the Dutch translated and adapted version of the Wechsler Adult Intelligence Scale (WAIS)*. Lisse: Swets and Zeitlinger.

van Beijsterveldt, C. E., Groen-Blokhuis, M., Hottenga, J. J., Franic, S., Hudziak, J. J., Lamb, D., . . . Schutte, N. (2013). The Young Netherlands Twin Register (YNTR): Longitudinal twin and family studies in over 70,000 children. *Twin Research and Human Genetics*, *16*, 252–267.

van der Sluis, S., Posthuma, D., & Dolan, C. V. (2013). TATES: Efficient multivariate genotype-phenotype analysis for genome-wide association studies. *PLoS Genetics*, *9*, e1003235.

van der Sluis, S., Verhage, M., Posthuma, D., & Dolan, C. V. (2010). Phenotypic complexity, measurement bias, and poor phenotypic resolution contribute to the missing heritability problem in genetic association studies. *PLoS One*, *5*, e13929.

van Dongen, J. P., Draisma, H. H. M., Martin, N. G., & Boomsma, D. I. (2012). The continuing value of twin studies in the omics era. *Nature Reviews Genetics*, *13*, 640–653.

Van Haasen, P., De Bruyn, E., Pijl, Y., Poortinga, Y., Lutje-Spelberg, H., Vander Steene, G., . . . Stinissen, J. (1986). *Wechsler intelligence scale for children-revised, Dutch Version*. Lisse, the Netherlands: Swets & Zetlinger BV.

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *The American Journal of Human Genetics*, *90*, 7–24.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale — Third edition, Dutch version*. Lisse, The Netherlands: Swets and Zeitlinger.

Wechsler, D., Kort, W., Compaan, E. L., Bleichrodt, N., Resing, W. C. M., & Schittkatte, M. (2002). *Wechsler Intelligence Scale for Children — Third edition. Dutch version*. Lisse, The Netherlands: Swets and Zeitlinger.