

Genetic and Environmental Contributions to Behavioral Stability and Change in Children 6–36 months of Age Using Louisville Twin Study Data

Deborah Winders Davis¹ · Deborah Finkel² · Eric Turkheimer³ · William Dickens⁴

Received: 31 July 2015 / Accepted: 7 October 2015 / Published online: 19 October 2015
© Springer Science+Business Media New York 2015

Abstract The Infant Behavior Record (IBR) from the Bayley Scales of Infant Development has been used to study behavioral development since the 1960s. Matheny (1983) examined behavioral development at 6, 12, 18, and 24 months from the Louisville Twin Study (LTS). The extracted temperament scales included *Task Orientation*, *Affect-Extraversion*, and *Activity*. He concluded that monozygotic twins were more similar than same-sex dizygotic twins on these dimensions. Since this seminal work was published, a larger LTS sample and more advanced analytical methods are available. In the current analyses, Cholesky decomposition was applied to behavioral data ($n = 1231$) from twins 6–36 months. Different patterns of genetic continuity vs genetic innovations were identified for each IBR scale. Single common genetic and shared environmental factors explained cross-age twin similarity in the *Activity* scale. Multiple shared environmental factors and a single genetic factor coming on line at age 18 months contributed to *Affect-Extraversion*. A single

shared environmental factor and multiple genetic factors explained cross-age twin similarity in *Task Orientation*.

Keywords Behavior · Childhood · Longitudinal · Twin studies · Louisville Twin Study

Introduction

Since the pioneering work of Thomas and Chess (1977), much has been written regarding the role of temperament in explaining individual differences in behavior (Auerbach et al. 2008; Bates et al. 1998; Rothbart and Derryberry 1981; Rothbart et al. 1994). These differences in behavioral expression have been found to relate to important outcomes across multiple domains (Blair and Diamond 2008; Booth-LaForce and Oxford 2008; Chatoor et al. 2000; Goldsmith et al. 2007; Graziano et al. 2010; Hager 2010; Martin and Holbrook 1985; Matheny 1989a; Matheny and Dolan 1980; Rothbart and Jones 1998; Rothbart and Hwang 2005; Rothbart et al. 2006). While differences exist in measurement and specific behavioral classifications, there is a growing body of literature suggesting that while temperament has a biological basis, behavioral expression is modifiable and the environment can mediate the relationship between temperament and a variety of outcomes (Comas et al. 2014; Jaffe et al. 2010; Rapee 2014; van Ijzendoorn and Bakermans-Kranenburg 2015; Wachs and Kohnstamm 2001). What is less clear, are the specific developmental mechanisms responsible for the changes in the expression of specific temperament characteristics over time. Important in understanding *how* the genes and environments interact, is also understanding the points in time when significant changes occur that might suggest a period of increased vulnerability or susceptibility to changes in the

Electronic supplementary material The online version of this article (doi:10.1007/s10519-015-9759-x) contains supplementary material, which is available to authorized users.

✉ Deborah Winders Davis
deborah.davis@louisville.edu

¹ Department of Pediatrics, University of Louisville School of Medicine, 571 S. Floyd Street, Suite 412, Louisville, KY, USA

² School of Social Sciences, Indiana University Southeast, New Albany, IN, USA

³ Department of Psychology, University of Virginia, Charlottesville, VA, USA

⁴ Department of Economics, Northeastern University, Boston, MA, USA

physical or psychosocial environments. Various temperament characteristics have been associated with mental and physical health, behavior, personality, and academic outcomes; to name a few (Blair and Razza 2007; Blair and Diamond 2008; Booth-LaForce and Oxford 2008; Gjone and Stevenson 1997; Goldsmith and Lemery 2000).

Previous research has shown both stability and change in children's individual behavioral expression across time and situation in early childhood (Matheny 1983, 1989b; Matheny et al. 1985; Saudino et al. 1996). Twin studies have been used to better understand the relative contribution of the genetic and environmental influences on stability and change in behavior (Matheny 1983; 1989b; Matheny et al. 1985; Saudino et al. 1996). In a classic paper by Matheny (1983), the Infant Behavior Record (IBR) from Bayley Scales of Infant Development (BSID; Bayley 1969) was used to assess behavioral stability in approximately 300 twin pairs from the Louisville Twin Study (LTS) at 6, 12, 18, and 24 months. In a previous study, Matheny (1980) reduced 25 items of the IBR to three over-arching temperament scales (task orientation, affect-extraversion, and activity) and two additional non-temperament dimensions (auditory-visual awareness and motor coordination). These three temperament scales were used in subsequent studies by Matheny (1983) and others in twin studies (Saudino and Cherny 2001; Saudino et al. 1996) to better understand the genetic influences on behavioral stability and change in early childhood. These early works made important contributions to the literature about the development of temperament-related behavior. For example, the ideas were just emerging that while, overall, for all developmental phenotypes there is a gradual progression in behavioral complexity; there are individual differences in the rates of expression and within individual children the unfolding occurs in episodes of acceleration and lags (Wilson 1983). Twin studies demonstrated that these accelerations and lags were more synchronous for monozygotic twin pairs (Matheny 1980, 1983, Wilson 1983; Wilson and Matheny 1983, 1986). Wilson (1983) recognized that there was an ongoing and dynamic interaction between genes and the environment. These prior studies from LTS data examined temperament-rated behavior using the IBR, parent-report, and laboratory observations.

Since Matheny's classic paper in 1983, the participants in the LTS more than doubled and were followed longer in the longitudinal study. In addition, since 1983, new analytical methods have been developed that allow for greater precision in the data analysis. Later, using newer analytical methods, a similar sample size ($n = 301$ twin pairs) from the McArthur Longitudinal Twin Study (MALTS), and the same methodology used by Matheny (1983); Saudino et al. (1996) examined the three behavioral scales from the IBR

at 14, 20, and 24 months of age and, in 2001, in the same sample (Saudino and Cherny 2001), extended their work to 36 months. A modified IBR was used at 36 months. The previous three factors were extracted at 14, 20, and 24 months. However, *Affect-Extraversion* did not emerge at 36 months. Saudino and Cherny found significant heritability across ages with new genetic effects on *Activity* at 20 months and *Task Orientation* at 36 months.

Braungart et al. (1992) also used the IBR to examine the three behavioral scales previously extracted by Matheny (1980) with a sample of non-adoptive ($n = 190$) and adoptive siblings ($n = 160$) from the Colorado Adoption Project, which they compared to data previously reported by Matheny (1983) on twins from the LTS ($n = 270$). They concluded that the heritability estimates were 35–57 % and that there was very little contribution from the shared environment.

It is clear from the literature that temperament has, at least, a moderate biological basis, but the behavioral expression of innate temperament characteristics and adverse outcomes associated with more negative aspects of temperament is modifiable by the environment. The LTS has the potential to replicate and extend important research due to the frequency and comprehensiveness (IBR, parent-report, and laboratory observation) of the data collection from infancy through adolescence with larger sample sizes to confirm existing findings and identify new patterns of genetic and environmental influences. To demonstrate the utility of revisiting the LTS data, the current study aims to replicate and extend the work done by Matheny (1983), Saudino et al. (1996), and Saudino and Cherny (2001) using the IBR data from the Louisville Twin Study that were collected at 6, 9, 12, 18, 24, 30, and 36 months of age in more than 600 pairs of twins. The current dataset represents a broader range of ages and a larger sample size than the previous studies from the LTS and the MALTS.

Methods

Participants

Twins enrolled in the LTS were recruited from the Board of Health birth certificate records of twin births for families residing in the metropolitan Louisville, Kentucky area at the time of the twins' birth. The LTS sample consisted of families who were recruited to represent the full range of socioeconomic status, race, and ethnic diversity that was within the Louisville metropolitan area at the time of recruitment. Approximately 80 % of the participants are European-American, 18 % are African-American, and the remaining 2 % are of mixed or Asian ancestry. Occupations of heads of households, converted to Duncan's scores for

socioeconomic status (Hollingshead 1975), represented the entire distribution of social class, with an average score on the 100-point scale of 46.89 (SD = 26.9), a score typical for middle-level clerical workers. Mean gestational age of the LTS sample is 37.18 weeks (SD = 2.6), which is just below the population mean of 40 weeks for single-born neonates. Special efforts were made to retain recruited families in the study, and less than 10 % of the sample withdrew from the longitudinal study during the first 3 years.

A total of 1231 children born between 1957 and 1997 had data on the variables of interest between 6 and 36 months of age for the current analysis. Forty-seven percent of the sample participated in at least 6 of the 7 waves of measurement; 73 % participated in 4 or more waves. Zygosity was determined by blood sera analysis made when the twins were 36 months of age or older as part of the LTS protocol. Number of children participating in each wave is presented in Table 1; both same-sex and opposite-sex pairs were included. Prior to 36 months, zygosity was determined by a consensus of staff ratings of the twins' similarity of physical appearance. When repeated observations were completed between 6 and 36 months, there was 98 % agreement between the ratings and the blood typing results (Wilson and Matheny 1986). Previous studies have demonstrated that assessments of physical similarity are 90–95 % accurate when compared to serological tests (Nichols and Bilbro 1966). Results of blood sera testing were available for 79 % of the sample. Forty-seven percent of pairs included here were monozygotic twins (MZ) and 53 % were dizygotic twins (DZ).

Measures

The twins were administered the Bayley Scales of Infant Development (Bayley 1969, 1993) individually by separate examiners at each visit to the study center at 6, 9, 12, 18,

24, 30, and 36 months of age (± 1 week). The testing schedule was arranged so that examiners did not test the same twin on successive visits. After they had been assessed by the mental and motor scales, procedures that lasted approximately 90 min, the Infant Behavior Record (IBR) was filled out in accord with the BSID manual. Following Matheny (1980, 1983) and Saudino et al. (1996), three scales were created from IBR items. Task orientation indicates the infant's general engagement with the testing task and includes items assessing object orientation (item 8), goal directedness (item 11), and attention span (item 12). Affect-extraversion measures the extent to which children engaged in the social dynamics of the testing situation and includes responsiveness to examiner (item 2), cooperativeness (item 4), fearfulness (item 5, reverse scored), and happiness (item 7). The activity scale focused on the general level of activity and energy and included activity (item 14), body motion (21), and energy level (item 25). Comparisons of factor structure across testing occasions indicated that the factor structure did not vary systematically with age. To avoid variance in measurement (cf. Wicherts et al. 2004), an invariant definition of scales at each testing occasion was created by standardizing the individual IBR items relative to the respective means and variances at 6 months. Items were standardized to a mean of 5 and a standard deviation of 1, and then summed to create the IBR scales. The inter-observer reliabilities, based on a subsample of 57 infants, were 0.82, 0.87, and 0.79 for task orientation, affect-extraversion, and activity, respectively (Matheny 1983). Sample sizes, mean scores (and standard deviations) for the IBR scales are presented in Table 1. Means and standard deviations were generally stable across age, but differed somewhat between boys and girls, especially at the later ages. For that reason, scales were sex-corrected before behavior genetic analyses were conducted.

Table 1 Mean and standard deviations for the three IBR factors at each wave

Wave	N M/F	Twin pairs MZ/DZ	Task orientation		Affect-extraversion		Activity	
			Boys	Girls	Boys	Girls	Boys	Girls
6 months	427/458	181/225	13.5 (4.1)	13.9 (4.0)	15.5 (2.7)	15.5 (2.6)	13.3 (2.8)	13.0 (2.6)
9 months	355/381	172/177	16.4 (3.3)	16.2 (3.2)	16.1 (2.7)	16.2 (2.9)	13.4 (2.6)	13.1 (2.6)
12 months	465/507	210/237	15.0 (3.8)	15.4 (3.6)	15.9 (3.0)	16.0 (3.0)	13.8 (2.8)	13.7 (2.5)
18 months	427/455	193/217	14.5 (4.1)	15.1 (3.6)	15.1 (3.3)	15.1 (3.3)	14.9 (3.4)	14.0 (2.8)
24 months	444/504	206/240	15.8 (3.6)	16.2 (3.9)	16.5 (3.6)	16.5 (3.5)	14.6 (3.1)	14.1 (2.9)
30 months	404/423	177/214	14.6 (3.7)	15.2 (3.3)	15.7 (3.6)	16.4 (3.2)	15.0 (3.3)	13.9 (3.0)
36 months	346/333	147/169	15.4 (3.8)	16.1 (3.4)	16.8 (3.3)	17.4 (3.4)	14.4 (2.8)	14.0 (2.8)

N indicates the number of males (M) and females (F)

Statistical model

The variance in any trait can be divided into three separate components: additive genetic effects (A), shared rearing environmental effects that serve to make the members of a family more similar (C), and nonshared environmental effects including error (E). By fitting structural equation models to the observed MZ and DZ covariance matrices, we can estimate the proportion of total variance accounted for by the variance in genetic factors, shared environment factors, and nonshared environment factors. Univariate analyses of the IBR scales were conducted separately at each age. Multivariate analyses provide additional power for estimating parameters when compared to univariate approaches (Schmitz et al. 1998); therefore a Cholesky decomposition of the genetic and environmental covariances was applied to model the consecutive waves of IBR testing. Figure 1 presents the Cholesky decomposition of the additive genetic factors: each latent genetic factor (A6 through A36) impacts performance at that age and at each subsequent age. In the full model, similar paths were included for shared environment (C6 through C36) and nonshared environmental (E6 through E36). To address issues of continuity and change in sources of variance, the focus of model testing was not on individual path coefficients but on the role of the latent factors (A, C, and E) at each age. The Cholesky model allows us to test hypotheses about single general sources of variance (A6, C6, and E6) versus multiple age-specific sources of variance (e.g. A9 through A36). Heritability estimates can increase throughout childhood as the result of the amplification of original genetic factors (A6) or from the emergence of new genetic sources of variance in childhood (e.g. A9 through A36). Similar hypotheses about shared and nonshared environmental influences can be examined. Hypotheses were tested by comparing model fit indices; nested models were compared using the likelihood ratio test (LRT) obtained by taking the difference between the obtained model fits (log likelihoods) and testing its significance with

the degrees of freedom equal to the difference in the number of parameters of the two models. Akaike's Information Criterion (AIC), which is the log-likelihood minus twice the degrees of freedom, was used to identify the most parsimonious model.

Results

Correlations

Phenotypic age-to-age correlations are presented in Table 2. The correlations were modest in the younger ages (range 0.13–0.30) but tended to be slightly larger at the later ages (range 0.23–0.41). Thus, the pattern of correlations indicates some continuity, and possibly increasing continuity, combined with substantial change in the rank order of individuals during childhood. Twin intraclass correlations on the three IBR scales, both within age and across age, are presented in Table 3. The patterns of correlations differed across IBR scales. For *Task Orientation* the pattern of MZ and DZ intraclass correlations within-age was generally consistent from one age to another and there were significant twin correlations across ages. In contrast, for *Affect-Extraversion* the difference between MZ and DZ correlations tended to increase with age, suggesting increases in genetic influences. Across ages, twin similarity tended to decrease with increasing distance between measurement occasions. The pattern of twin correlations for the *Activity* scale was inconsistent and the MZ-DZ differences across the first 6 ages showed no clear pattern of increasing or decreasing similarity.

Univariate analyses

As a preliminary step to multivariate analyses, univariate analyses of twin similarity were conducted separately at each age for each IBR scale. Components of variance estimated by the full model are presented in Table 4;

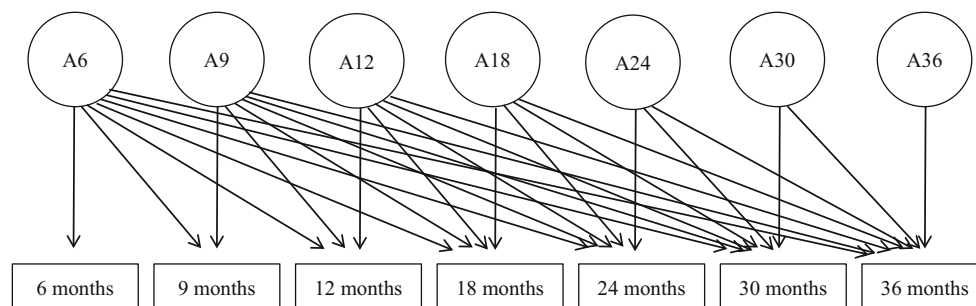


Fig. 1 Cholesky model of latent additive genetic factors influence behavior at ages 6 through 36 months. The full model included C6 through C36 and E6 through E36

Table 2 Phenotypic stability of IBR temperament factors across age

	6–9 months	9–12 months	12–18 months	18–24 months	24–30 months	30–36 months
Task orientation	0.17**	0.26**	0.26**	0.32**	0.34**	0.40**
Affect-extraversion	0.13**	0.23**	0.13**	0.23**	0.28**	0.23**
Activity	0.29**	0.24**	0.30**	0.37**	0.36**	0.34**

** $p < 0.01$ **Table 3** Twin intraclass correlations, within age and across age

Age	6 months		9 months		12 months		18 months		24 months		30 months		36 months	
	MZ	DZ	MZ	DZ	MZ	DZ	MZ	DZ	MZ	DZ	MZ	DZ	MZ	DZ
<i>Task Orientation</i>														
6	0.45**	0.36**												
9	0.26**	0.17**	0.32**	0.25**										
12	0.15**	0.18**	0.33**	0.16**	0.40**	0.33**								
18	0.07	0.16**	0.09	0.11	0.21**	0.23**	0.37**	0.30**						
24	0.17**	0.29**	0.12	0.11	0.03	0.26**	0.27**	0.34**	0.30**	0.38**				
30	0.15*	0.15*	0.21**	0.15*	0.07	0.14**	0.20**	0.25**	0.24**	0.23**	0.41**	0.28**		
36	0.18*	0.32**	0.18*	0.26**	0.04	0.23**	0.17**	0.35**	0.31**	0.33**	0.47**	0.31**	0.62**	0.39**
<i>Affect-Extraversion</i>														
6	0.24**	0.24**												
9	0.16**	0.22**	0.11	0.19**										
12	0.05	0.10	0.29**	0.20**	0.27**	0.26**								
18	0.10	0.08	0.09	0.10	0.12*	0.17**	0.28**	0.12*						
24	0.06	0.10	0.06	0.13	0.16**	0.09	0.33**	0.17**	0.40**	0.24**				
30	0.07	0.03	-0.01	0.02	0.07	0.01	0.30**	0.04	0.25**	0.07	0.41**	0.11		
36	0.07	0.11	0.05	-0.02	-0.15	0.03	0.19**	0.09	0.20**	0.05	0.28**	0.11	0.41**	0.18**
<i>Activity</i>														
6	0.43**	0.21**												
9	0.28**	0.15*	0.35**	0.27**										
12	0.21**	0.15*	0.26**	0.23**	0.32**	0.27**								
18	0.13*	0.10	0.16**	0.21**	0.20**	0.21**	0.44**	0.29**						
24	0.09	0.07	0.07	0.13*	0.16**	0.26**	0.34**	0.33**	0.33**	0.29**				
30	0.08	0.03	0.09	0.13	0.18**	0.19**	0.35**	0.17**	0.29**	0.29**	0.48**	0.27**		
36	-0.01	-0.03	0.05	0.06	0.08	0.19**	0.12	0.14	0.20**	0.12	0.35**	0.20**	0.29**	0.28**

* $p < 0.05$; ** $p < 0.01$

significance indicates parameter estimates in the full model with 95 % confidence intervals that did not contain zero. As expected from pattern of twin correlations, heritability estimates for *Affect Extraversion* were low at earlier ages and higher at later ages. Estimates of shared environmental components of variance demonstrated the opposite pattern. For both *Task Orientation* and *Activity*, all three components of variance were fairly stable, with some fairly small values for heritability at age 24 months for *Task Orientation* and 36 months for *Activity*. For all three scales,

nonshared environmental influence is consistently significant and strong.

Multivariate analyses

The focus of model comparison was to test hypotheses about continuity and change in genetic and environmental factors influencing IBR factors over age. A model assuming complete continuity in twin similarity would include single common factors: A6 and C6. Results of testing

Table 4 Results of univariate twin analysis

	6 months	9 months	12 months	18 months	24 months	30 months	36 months
<i>Task orientation</i>							
Heritability	0.19	0.17	0.27	0.22	0.00	0.26	0.48*
Shared environment	0.27	0.16	0.18	0.18	0.35*	0.15	0.14
Nonshared environment	0.54*	0.67*	0.56*	0.61*	0.65*	0.59*	0.38*
<i>Affect Extraversion</i>							
Heritability	0.03	0.00	0.08	0.26	0.29	0.37*	0.44*
Shared environment	0.23	0.16	0.21	0.01	0.10	0.00	0.00
Nonshared environment	0.74*	0.84*	0.71*	0.73*	0.60*	0.63*	0.56*
<i>Activity</i>							
Heritability	0.44*	0.20	0.23	0.50*	0.17	0.50*	0.05
Shared environment	0.00	0.17	0.14	0.01	0.19	0.00	0.25
Nonshared environment	0.56*	0.63*	0.63*	0.49*	0.64*	0.49*	0.69

* $p < 0.05$

versions of this model for *Activity* are reported in Table 5. First, the full model estimating all parameters was fit to the data and served as the baseline for model comparisons. In model 2, latent genetic factors A9 through A36 were dropped from the model, retaining only A6. Fit of model 2 did not differ significantly from the baseline model (LRT = 19, $df = 21$, $p = 0.599$), indicating A9 through A12 were not necessary to adequately model twin similarity in *Activity*. In model 3, latent shared environmental factors C9 through C36 were dropped from the model, retaining only C6. Again, fit of the model did not differ significantly from the baseline model (LRT = 2, $df = 21$, $p = 0.999$). Model 4 represents a combination of models 2 and 3 in which only A6 and C6 were retained; fit of the model did not differ significantly from the baseline model (LRT = 58, $df = 42$, $p = 0.053$). Moreover, the difference between models 2 and 4 was significant (LRT = 38, $df = 21$, $p = 0.013$) indicating that adding C6 to model 2 produced a significant improvement in fit. Similarly, the difference between models 3 and 4 (LRT = 55, $df = 21$, $p < 0.001$) indicated that adding A6 to model 3 significantly improved model fit. In the final model (model 5), all

E contributions to cross-age similarity were dropped, which resulted in a smaller value for AIC. Other models were tested, but none improved on model fit. Thus, a single common genetic factor (A6) and a single common shared environmental factor (C6) were sufficient to explain cross-age twin similarity in *Activity*. Figure 2 presents the path diagram of model 5 (solid lines indicate significant path coefficients and dashed lines indicated nonsignificant path coefficients). The associated path coefficients and confidence intervals are reported in Supplemental Table 1. Because the focus of model fitting was on continuity and change as indicated by the latent factors (A, C, and E), the final models may contain individual nonsignificant paths. Standardized variance components at each age as estimated by model 5 are presented in the top panel of Fig. 5. Heritability decreased somewhat with age and shared environmental factors were more important starting at age 18 months.

Initially, the pattern of model fitting for *Affect-Extraversion* was similar to the results for *Activity*; results are presented in Table 6. Comparing the fit of models 2 through 4 to the baseline model again suggested single

Table 5 Model-fitting results for activity

Model	–2LL	df	Parms	AIC	LRT	df	p value
1. Full model	26,674	5490	91	15,694			
2. Keep only A6	26,693	5511	70	15,671	19	21	0.599
3. Keep only C6	26,676	5511	70	15,654	2	21	0.999
4. Combine models 2 & 3	26,731	5532	49	15,668	58	42	0.063
5. Model 4 + only age specific E	26,755	5553	28	15,665	97	81	0.065

–2LL is the log-likelihood indicating model fit; Parmes indicates the number of parameters fit by the model; AIC is Akaike's Information Criterion; LRT is the likelihood ratio test comparing the fit of each model to the full model

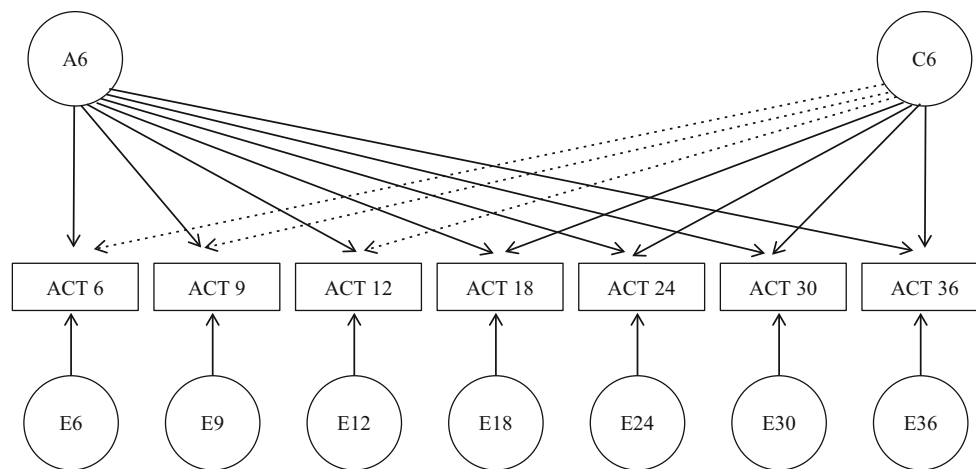


Fig. 2 Path diagram representing results for the best-fitting model for Activity: *solid lines* indicate significant path coefficients and *dashed lines* indicated nonsignificant path coefficients. The associated path coefficients and confidence intervals are reported in Supplemental Table 1

Table 6 Model-fitting results for affect extraversion

Model	–2LL	df	Parms	AIC	LRT	df	p value
1. Full model	28,041	5490	91	17,061			
2. Keep only A6	28,048	5511	70	17,026	6	21	0.999
3. Keep only C6	28,045	5511	70	17,022	3	21	0.999
4. Combine models 2 and 3	28,081	5532	49	17,017	40	42	0.573
5. Drop A6, A9, A12, C18, C24, C30, C36	28,048	5518	63	17,012	6	28	0.999
6. Model 5 + drop A24, A30, A36	28,052	5524	57	17,004	11	34	0.999
7. Model 5 + drop C9	28,062	5524	57	17,014	21	34	0.962
8. Model 5 + drop C12	28,048	5523	58	17,002	7	33	0.999
9. Keep only C6, C9, A18	28,060	5529	52	17,002	19	39	0.997
10. Model 9 + only age-specific E	28,092	5550	31	16,992	51	60	0.792

–2LL is the log-likelihood indicating model fit; Parms indicates the number of parameters fit by the model; AIC is Akaike's Information Criterion; LRT is the likelihood ratio test comparing the fit of each model to the full model

common A6 and C6 factors. However, the cross-age correlations in Table 3 and the univariate results in Table 4 suggested a more complicated pattern of genetic and environmental contributions to cross-age similarity; therefore, additional models were tested. Building on the univariate results, model 5 included C and E contributions up to age 12 months and A and E contributions from 18 months and older. Fit for model 5 differed by only 6 from the baseline model ($df = 28, p = 0.999$) and the AIC for model 5 was smaller than models 1 through 4. Model 6 tested whether a single source of genetic variance (A18) was sufficient to explain the data; results indicated that model 6 did not differ significantly from model 5 (LRT = 4, $df = 6, p = 0.637$) or the baseline model. Next, the role of C9 and C12 were tested independently in models 7 and 8, respectively. The difference between models 5 and 7, testing the role of C9, was significant (LRT = 14, $df = 6, p = 0.025$), but the difference

between models 5 and 8 was not significant (LRT = 1, $df = 5, p = 0.987$). Combining these model results suggested that the best-fitting model would include C6, C9, and A18 (model 9). AIC values reported in Table 6 indicate that model 9 fit better than all previous models. In the final model (model 10), all E contributions to cross-age similarity were dropped. This model resulted in a smaller value for AIC and did not differ significantly from the baseline model; thus model 10 represents the model with the fewest latent factors (A, C, or E) that adequately models cross-age similarity in *Affect-Extraversion*. Figure 3 presents the path diagram of model 10. The associated path coefficients and confidence intervals are reported in Supplemental Table 2. The path diagram indicates continuity (C6) and change (C9) in shared environmental influences and continuity in genetic influences that do not begin to impact behavior until age 18 months (A18). Standardized variance components at each age as estimated

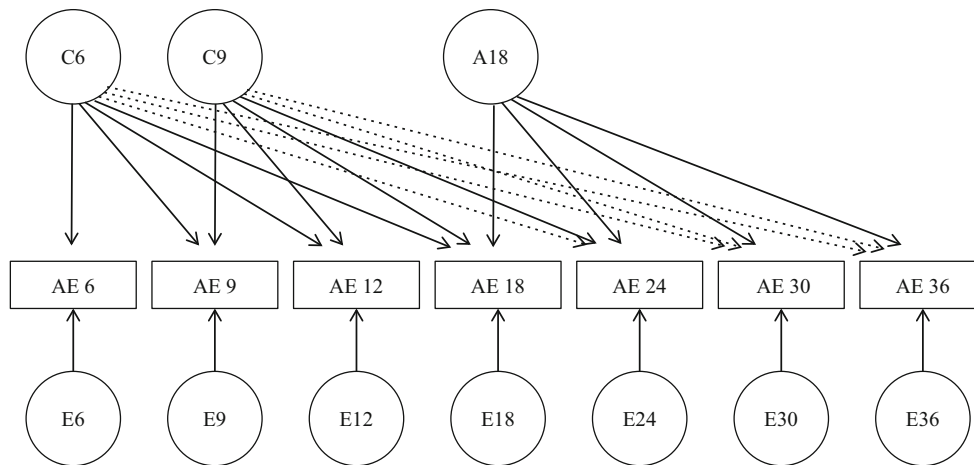


Fig. 3 Path diagram representing results for the best-fitting model for Affect Extraversion: *solid lines* indicate significant path coefficients and *dashed lines* indicated nonsignificant path coefficients. The

associated path coefficients and confidence intervals are reported in Supplemental Table 2

by model 10 are presented in the middle panel of Fig. 5. As expected from univariate results, twin similarity up to age 12 months resulted from shared environmental factors; at age 18 months and older, twin similarity resulted primarily from genetic factors.

Finally, results of model comparison for *Task Orientation* are presented in Table 7. We tested whether a single common factor (A6 or C6) could explain cross-age similarity. Neither model 2 or 3 represented a significant change in model fit from the baseline model, so the two models were combined in model 4, but model fit was significantly reduced compared with the baseline model (LRT = 85, $df = 42$, $p < 0.001$). In other words, the model was reduced too far to adequately fit the data. In models 5 through 10, the impact of adding back each latent

genetic factor (A9 through A36) was tested independently. Results indicated that model 5 (adding A9) and model 6 (adding A12) did not differ significantly from the baseline model and these models had lower AIC values than previous models. Models 7 through 10, however, did not indicate any improvement in model fit. Combining these model results suggested that the best-fitting model would include C6, A6, A9, and A12 (model 11). The difference between model 11 and the baseline model was minimal (LRT = 6, $df = 31$, $p = 0.999$). Moreover, dropping any individual latent factor from model 11 resulted in a significant decrease in model fit: drop C6 (LRT = 51, $df = 7$, $p < 0.001$), drop A6 (LRT = 36, $df = 7$, $p < 0.001$), drop A9 (LRT = 33, $df = 6$, $p < 0.001$), and drop A12 (LRT = 32, $df = 5$, $p < 0.001$). In the final model (model

Table 7 Model-fitting results for task orientation

Model	-2LL	df	Parms	AIC	LRT	df	p value
1. Full model	29,125	5495	91	18,136			
2. Keep only A6	29,138	5515	70	18,108	14	21	0.880
3. Keep only C6	29,128	5515	70	18,098	3	21	0.999
4. Combine models 2 and 3	20,210	5536	49	18,138	85	42	0.000
5. Model 4 + A9	29,162	5530	55	18,102	38	36	0.387
6. Model 4 + A12	29,163	5531	54	18,101	39	37	0.388
7. Model 4 + A18	29,176	5532	53	18,112	51	38	0.073
8. Model 4 + A24	29,177	5533	53	18,111	52	39	0.075
9. Model 4 + A30	29,179	5534	51	18,111	55	40	0.059
10. Model 4 + A36	29,206	5535	50	18,136	81	41	0.000
11. Keep only C6, A6, A9, A12	29,130	5525	60	18,080	6	31	0.999
12. Model 11 + only age-specific E	29,172	5546	39	18,080	48	52	0.653

-2LL is the log-likelihood indicating model fit; Parmns indicates the number of parameters fit by the model; AIC is Akaike’s Information Criterion; LRT is the likelihood ratio test comparing the fit of each model to the full model

12), all E contributions to cross-age similarity were dropped. This model resulted in a similar value for AIC and did not differ significantly from the baseline model. Thus, model 12 represents the model with the fewest latent factors that adequately models cross-age similarity in *Task Orientation*. Figure 4 presents the path diagram of model 12. The associated path coefficients and confidence intervals are reported in Supplemental Table 3. The path diagram presents continuity in shared environmental influences on *Task Orientation* and new genetic variance coming on line at ages 9 and 12 months. Standardized variance components at each age as estimated by model 12 are presented in the bottom panel of Fig. 5. Heritability varies across age, but estimates represent a fairly good match to univariate estimates reported in Table 4.

Discussion

The current study replicates and extends previous longitudinal studies of temperament using scales (*Task Orientation*, *Affect-Extraversion*, and *Activity*) derived from the IBR at ages 6, 9, 12, 18, 24, 30, and 36 months of age with more than 600 pairs of twins from the Louisville Twin Study. The current study extended that of Matheny (1983) by using a larger sample size and testing ages of 9, 30, and 36 months in addition to the 6, 12, 18, and 24 months previously reported. With a similar sample size to that of Saudino and Cherny (2001), we examined younger children and more frequent testing points.

Age-to-age correlations were modest in the younger ages and somewhat larger at the later ages suggesting some

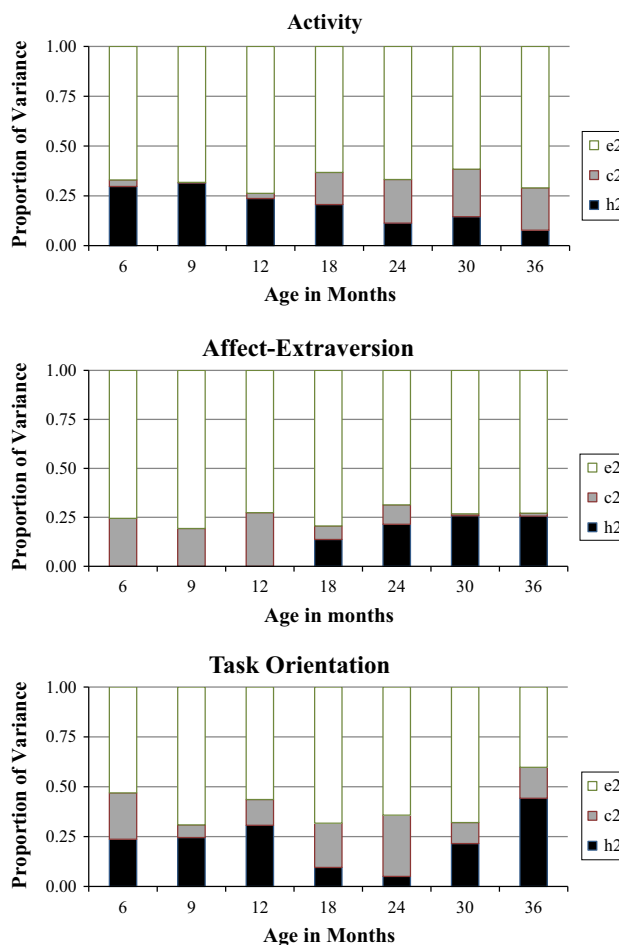


Fig. 5 Standardized variance components estimated from the best-fitting model for each IBR factor: h2 = heritability, c2 = shared environment, e2 = nonshared environment

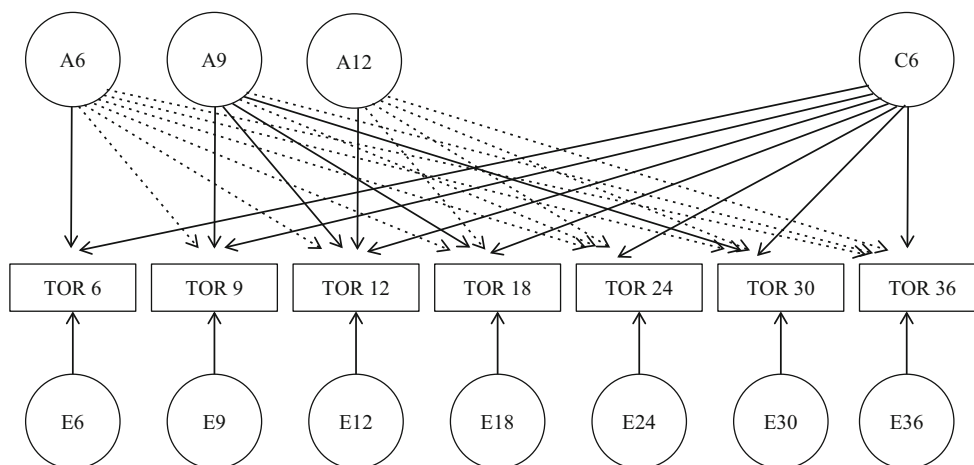


Fig. 4 Path diagram representing results for the best-fitting model for *Task Orientation*: solid lines indicate significant path coefficients and dashed lines indicated nonsignificant path coefficients. The associated

path coefficients and confidence intervals are reported in Supplemental Table 3

continuity, and possibly increasing continuity, combined with substantial change in the rank order of individuals during childhood. Multivariate analyses provided evidence for both continuity and discontinuity in sources of variance in the IBR scales. Across studies, the phenotypic stability correlations showed a similar pattern, but the correlations in the current study were higher than those reported by Saudino and Cherny (2001) and lower than those reported by Matheny (1983). Cross-age interclass correlations were similar across studies. Consistent with the literature, our findings support the notion that temperament, in general, is influenced by genetics to some degree, but behavioral expression is also attributable to environmental influences, which varies by dimension.

For the *Activity* scale, for example, a single common genetic factor (A6) and a single common shared environmental factor (C6) explained cross-age twin similarity, emphasizing continuity in genetic and shared environmental sources of variance for this scale. As with all three of the IBR scales, nonshared environmental influences on *Activity* were unique to each age of measurement and, thus, were a source of discontinuity across age. While Saudino and Cherny (2001) found a common genetic factor across 14, 20, 24, and 36 months; they also found new genetic variance at 36 months, which differs from our findings. Also, in contrast to our findings, Saudino and Cherny found no shared environmental influence while our findings suggest a single common shared environmental effect (C6). Consistent with Saudino and Cherny (2001), we found unique nonshared environmental influences at each age. More frequent data points across greater ages (6–36 months) may contribute to the differences in findings.

In contrast to the results for *Activity*, evidence for discontinuity in sources of genetic and shared environmental variance was found for *Affect-Extraversion*: before age 18 months the twin similarity was entirely attributable to shared environmental factors (C6 and C9), whereas at 18 months, the influence of shared environment began to wane and the influence of genetic factors (A18) became predominant. In other words, although a single genetic factor was identified for *Affect-Extraversion*, it did not emerge until age 18 months. Saudino and Cherny (2001) found a common genetic factor that emerged at 14 months and new genetic variance at 20 and 24 months. Comparisons between the studies are difficult since they did not examine children younger than 14 months or beyond 24 months on this dimension. In addition, the current study demonstrated new shared environmental effects at 6 and 9 months and unique nonshared environmental effects at each age. Saudino and Cherny (2001) demonstrated a similar effect of the nonshared environment, but, in contrast to the current study, they found no shared

environmental influence. This finding may reflect data collection at a period when language is beginning to emerge and parent responsivity to early communication skills may influence the expression of the *Affect-Extraversion* dimension of temperament.

Finally, *Task Orientation* demonstrated yet another pattern of continuity. All genetic sources of variance emerged early in childhood at ages 6, 9, and 12 months and continued to influence behavior thereafter. A single common factor explained all shared environmental variance for *Task Orientation*. Thus, although new genetic variance emerged at 9 and 12 months, the primary pattern for *Task Orientation* is one of continuity of genetic and shared environmental influences. There is the possibility of changes in the nature of genetic influence between 18 and 30 months. Saudino and Cherny (2001) demonstrated a similar pattern of genetic continuity across 14–36 months. However, unlike the current study, they found new genetic contributions at 36 months, but questioned whether the difference might be attributable to differences in the factor structure at 36 months. Unlike the current study, others found no influence of the shared environment (Braungart et al. 1992; Saudino and Cherny 2001), but, found a similar effect of the nonshared environment (Braungart et al. 1992; Saudino and Cherny 2001).

Understanding the factors associated with stability and change over infancy and early childhood is important as it is a period of relatively rapid change and neurobiological differentiation (Couperus and Nelson 2006; Sameroff and Fiese 2000; Shonkoff 2010; Shonkoff and Marshall 2000). Current developmental theories posit that developmental processes are dynamic and transactional across all domains (Sameroff and Fiese 2000; Shonkoff 2010; Thelen and Smith 1996), including temperament and behavior. Transition periods are important to understanding how genes and environments interact to produce differential outcomes (Thelen and Smith 1996) because it is at these points of transition that the system is most receptive or vulnerable to changes in the environmental context (Sameroff and Fiese 2000; Shonkoff 2010; Thelen and Smith 1996).

Our findings confirm that of others in that there is evidence a biological component to the expression of these three temperament factors, but there is a consistent influence from the shared and nonshared environment that varies somewhat across factors. More data are needed to better understand the specific environmental variations that contribute to differences in behavior and how these environmental characteristics interact at periods of change in genetics expression. This is particularly important with temperament as it has been shown that while there is a genetic basis for the expression of temperament, outcomes are modifiable by the environment. It has been shown that children with more negative temperament characteristics

are at greater risk for adverse outcomes (Eisenberg et al. 2004; Rothbart 2011). However, when the physical and psychosocial environments are more flexible and child-focused, the negative outcomes associated with more difficult temperament styles can be moderated (Eisenberg et al. 2005; Keogh 2003; Rothbart 2011). This is consistent with the “goodness-of-fit” concept first described by Thomas and Chess (1977). It seems likely that interventions aimed at providing a supportive environment, especially during the key transition periods, would be optimal as the system is reorganizing during those periods. Children with more difficult temperament styles may need more sensitive and structured environments to promote optimal developmental outcomes, especially as it relates to self-regulation (Eisenberg et al. 2004, 2005; Rothbart 2011; Rothbart et al. 2000a, 2011).

Future studies should include longitudinal analyses beyond early childhood to extend into middle childhood, adolescence, and adulthood to fully appreciate the fluctuations and stabilities over time. Prospective data collection into adulthood can create a unique opportunity to explore an infancy to middle adulthood model of genetic-environmental influence on behavioral expression and a clearer understanding of the complex relationships responsible for adverse outcomes as well as scales such as resiliency and adaptation. A greater understanding is needed of the shared and nonshared environmental factors associated with changes in the behavioral expression of specific temperament characteristics over time to promote optimal health and developmental outcomes. Additionally, future studies should examine temperament stability and change and associations with various outcomes using newer temperament dimensions (Effortful Control, Negative Affect, and Extroversion) such as those described by Rothbart and co-authors (2000b, 2001a, b, 2011). The childhood temperament dimensions have been described as precursors for adult personality (Rothbart and Ahadi 1994; Rothbart et al. 2000a). New data collection at the LTS should include such personality measures to allow for life course analyses.

Acknowledgments We are grateful to the original Louisville Twin Study researchers and staff for the meticulous work they did to make it possible for us to revisit these data. This research was supported, in part, by a grant from the National Institute of Aging (1 R03 AG048850-01) and by the Department of Pediatrics at the University of Louisville.

Compliance with Ethical Standards

Conflict of Interest Deborah Winders Davis, Deborah Finkel, Eric Turkheimer, and William Dickens declare that they have no conflict of interest.

Human Rights and Informed Consent The secondary data analysis was approved by the Institution Review Board at the University of Louisville and all procedures were done in accordance with their

ethical standards. Informed consent was obtained from all participants for the original data collection.

References

- Auerbach JG, Berger A, Atzaba-Poria N, Arbelle S, Cypin N, Friedman A, Landau R (2008) Temperament at 7, 12, and 25 months in children at familial risk for ADHD. *Infant Child Dev* 17:321–338
- Bates JE, Dodge KA, Pettit GS, Ridge B (1998) Interaction of temperamental resistance to control and restrictive parenting in the development of existing behavior. *Dev Psychol* 34:982–995
- Bayley N (1969) Bayley scales of infant development. The Psychological Corporation, San Antonio
- Bayley N (1993) Bayley scales of infant development. The Psychological Corporation, San Antonio
- Blair C, Diamond A (2008) Biological processes in prevention and intervention: the promotion of self-regulation as a means of preventing school failure. *Dev Psychopathol* 20:899–911
- Blair C, Razza RP (2007) Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Dev* 78:647–663
- Booth-LaForce C, Oxford ML (2008) Trajectories of social withdrawal from grades 1 to 6: prediction from early parenting, attachment, and temperament. *Dev Psychol* 44:1298–1313
- Braungart JM, Plomin R, DeFries JC, Fulker DW (1992) Genetic influence on tester-rated infant temperament as assessed by Bayley’s Infant Behavior Record: nonadoptive and adoptive siblings and twins. *Dev Psychol* 28:40–47
- Chatoor I, Ganiban J, Hirsch R, Borman-Spurrell E, Mrazek DA (2000) Maternal characteristics and toddler temperament in infantile anorexia. *J Am Acad Child Adolesc Psychiatry* 39:743–751
- Comas M, Valentino K, Bridgett D, Hayden L (2014) The direct and interactive effects of physical abuse severity and negative affectivity on length of psychiatric hospitalization: evidence of differential reactivity to adverse environments in psychiatrically high-risk youth. *Child Psychiatry Hum Dev* 45:220–228
- Couperus JW, Nelson CA (2006) Early brain development and plasticity. In: McCarthey K, Phillips D (eds) *Blackwell handbook of early childhood development*. Blackwell, Malden, pp 85–105
- Eisenberg N, Spinrad TL, Fabes RA, Reiser M, Cumberland A, Shepard S, Valiente C, Losoya SH, Guthrie IK, Thompson M (2004) The relations of effortful control and impulsivity to children’s resiliency and adjustment. *Child Dev* 75:25–46
- Eisenberg N, Zhou Q, Spinrad TL, Valiente C, Fabes RA, Liew J (2005) Relations among positive parenting, children’s effortful control, and externalizing problems: A three-wave longitudinal study. *Child Dev* 76:1055–1071
- Gjone H, Stevenson J (1997) A longitudinal twin study of temperament and behavior problems: common genetic or environmental influences? *J Am Acad Child Adolesc Psychiatry* 36:1448–1456
- Goldsmith HH, Lemery KS (2000) Linking temperamental fearfulness and anxiety symptoms: a behavior-genetic perspective. *Biol Psychiatry* 48:1199–1209
- Goldsmith HH, Lemery-Chalfant K, Schmidt NL, Arneson CL, Schmidt CK (2007) Longitudinal analyses of affect, temperament, and childhood psychopathology. *Twin Res Hum Genet* 10:118–126
- Graziano PA, Calkins SD, Keane SP (2010) Toddler self-regulation skills predict risk for pediatric obesity. *Int J Obes* 34:633–641

- Hagger MS (2010) Sleep, self-regulation, self-control and health. *Stress Health* 26:181–185
- Hollingshead A (1975) Four Factor Index of social status. Dissertation in Sociology. Yale University, New Haven
- Jaffe M, Gullone E, Hughes EK (2010) The roles of temperamental dispositions and perceived parenting behaviours in the use of two emotion regulation strategies in late childhood. *J Appl Dev Psychol* 31:47–59
- Keogh B (2003) Temperament in the classroom: understanding individual differences. Brookes, Baltimore
- Martin RP, Holbrook J (1985) Relationship of temperament characteristics to the academic achievement of first-grade children. *J Psychoeduc Assess* 3:131–140
- Matheny AP Jr (1980) Bayley's infant behavior record: behavioral components and twin analyses. *Child Dev* 51:1157–1167
- Matheny AP Jr (1983) A longitudinal twin study of stability of components from Bayley's infant behavior record. *Child Dev* 54:356
- Matheny AP Jr (1989a) Injury prevention and temperament. In: Carey B, McDevitt S (eds) *Clinical and educational applications of temperament research*. Swets & Zeitlinger, Amsterdam, pp 103–106
- Matheny AP Jr (1989b) Children's behavioral inhibition over age and across situations: genetic similarity for a trait during change. *J Personal* 57:215–235
- Matheny AP Jr, Dolan A (1980) A twin study of personality and temperament during middle childhood. *J Res Personal* 14:224–234
- Matheny AP Jr, Wilson RS, Nuss S (1985) Toddler temperament: stability across settings and over ages. In: Chess S, Thomas A (eds) *Annual progress in child psychiatry and child development*. Brunner/Mazel, New York
- Nichols RC, Bilbro WC Jr (1966) The diagnosis of twin zygosity. *Hum Hered* 16:265–275
- Rapee RM (2014) Preschool environment and temperament as predictors of social and nonsocial anxiety disorders in middle adolescence. *J Am Acad Child Adolesc Psychiatry* 53:320–328
- Rothbart MK (2011) *Becoming who we are: temperament and personality in development*. Guilford Press, New York
- Rothbart MK, Ahadi SA (1994) Temperament and the development of personality. *J Abnorm Psychol* 103:55–66
- Rothbart MK, Derryberry D (1981) Development of individual differences in temperament. In: Lamb ME, Brown AL (eds) *Advances in developmental psychology*, vol 1. Erlbaum, Hillsdale, pp 37–86
- Rothbart MK, Hwang J (2005) Temperament and the development of competence and motivation. In: Elliot AJ, Dweck CS (eds) *Handbook of competence and motivation*. Guilford Press, New York, pp 167–184
- Rothbart MK, Jones LB (1998) Temperament, self-regulation, and education. *Sch Psychol Rev* 27:479–491
- Rothbart MK, Derryberry D, Posner M (1994) A psychobiological approach to the development of temperament. In: Bates JE, Wachs TD (eds) *Temperament: Individual differences at the interface of biology and behavior*. Am Psychol Assoc, Washington, DC
- Rothbart MK, Ahadi SA, Evans DE (2000a) Temperament and personality: origins and outcomes. *J Personal Soc Psychol* 78:122–135
- Rothbart MK, Derryberry D, Reed MA, Hershey K (2000b) Stability of temperament in childhood: laboratory infant assessment to parent report at seven years. In: Molfese DL, Molfese VJ (eds) *Temperament and personality development across the life span*. Erlbaum, Hillsdale
- Rothbart MK, Chew KH, Gartstein MA (2001a) Assessment of temperament in early development. In: Singer LT, Zeskind PS (eds) *Biobehavioral assessment of the infant*. Guilford Press, New York, pp 190–208
- Rothbart MK, Ahadi SA, Hershey KL, Fisher P (2001b) Investigations of temperament at three to seven years: the children's behavior questionnaire. *Child Dev* 72:1394–1408
- Rothbart MK, Posner MI, Kieras J (2006) Temperament, attention, and the development of self-regulation. In: McCartney K, Phillips D (eds) *Blackwell handbook of early childhood development*. Blackwell, Malden, pp 338–357
- Rothbart MK, Ellis LK, Posner M (2011) Temperament and self-regulation. In: Vohs KD, Baumeister RF (eds) *Handbook of self-regulation*. Guilford Press, New York, pp 441–460
- Sameroff AJ, Fiese BH (2000) Transactional regulation: the developmental ecology of early intervention. In: Shonkoff JP, Meisels SJ (eds) *Handbook of early intervention*. Cambridge University Press, Cambridge, pp 135–159
- Saudino KJ, Cherny SS (2001) Sources of continuity and change in observed temperament. In: Emde RN, Hewitt JK (eds) *Infancy to early childhood: genetic and environmental influences on developmental change*. Oxford University Press, New York, pp 89–110
- Saudino KJ, Plomin R, DeFries JC (1996) Tester-rated temperament at 14, 20 and 24 months: environmental change and genetic continuity. *Br J Dev Psychol* 14:129–144
- Schmitz S, Cherny SS, Fulker DF (1998) Increase in power through multivariate analyses. *Behav Genet* 28:357–363
- Shonkoff JP (2010) Building a new biodevelopmental framework to guide the future of early childhood policy. *Child Dev* 81:357–367
- Shonkoff JP, Marshall PC (2000) The biology of developmental vulnerability. In: Shonkoff JP, Meisels SJ (eds) *Handbook of early childhood intervention*. Cambridge University Press, New York, pp 35–53
- Thelen E, Smith L (1996) *A dynamic systems approach to the development of cognition and action*. MIT Press, Cambridge
- Thomas A, Chess S (1977) *Temperament and development*. Brunner/Mazel, New York
- van Ijzendoorn MH, Bakermans-Kranenburg MJ (2015) Genetic differential susceptibility on trial: meta-analytic support from randomized controlled experiments. *Dev Psychopathol* 27:151–162
- Wachs TD, Kohnstamm GA (2001) The bidirectional nature of temperamental-context links. In: Wachs TD, Kohnstamm GA (eds) *Temperament in context*. Erlbaum, Mahwah, pp 201–222
- Wicherts JM, Dolan CV, Hessen DJ, Oosterveld P, van Baal GCM, Boomsma DI, Span MM (2004) Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence* 32:509–537
- Wilson RS (1983) The Louisville Twin Study: developmental synchronies in behavior. *Child Dev* 54:298–316
- Wilson RS, Matheny AP Jr (1983) Assessment of temperament in infant twins. *Dev Psychol* 19:172–183
- Wilson RS, Matheny AP Jr (1986) Behavior-genetics research in infant temperament: the Louisville Twin Study. In: Plomin R, Dunn J (eds) *The study of temperament: changes, continuities and challenges*. Erlbaum, Hillsdale, pp 81–97