




Examining the factor structure and discriminative utility of the Infant Behavior Questionnaire–Revised in infant siblings of autistic children

Sooyeon Sung¹  | Angela Fenoglio¹ | Jason J. Wolff¹  | Robert T. Schultz² | Kelly N. Botteron³ | Stephen R. Dager⁴ | Annette M. Estes⁴ | Heather C. Hazlett⁵ | Lonnie Zwaigenbaum⁶ | Joseph Piven⁵ | Jed T. Elison¹  | for the IBIS Network

¹University of Minnesota, Minneapolis, Minnesota, USA

²Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA

³Washington University School of Medicine, Saint Louis, Missouri, USA

⁴University of Washington, Seattle, Washington, USA

⁵University of North Carolina—Chapel Hill, Chapel Hill, North Carolina, USA

⁶University of Alberta, Edmonton, Alberta, USA

Correspondence

Jed T. Elison, University of Minnesota, 51 East River Rd, Minneapolis, MN 55455, USA.

Email: jtelison@umn.edu

Funding information

Intellectual and Developmental Disabilities Research Center, Grant/Award Number: HD079124, HD087011, HD86984 and HD083091; NIH Autism Centers of Excellence Network, Grant/Award Number: R01 HD055741; Autism Speaks, Grant/Award Number: 6020; Simons Foundation, Grant/Award Number: 140209; NIH Research Career Development Awards, Grant/Award Number: K01 MH101653; NIH Research Project Grant Program, Grant/Award Number: R01 MH104324 and R01 MH116961

Abstract

Using the Infant Behavior Questionnaire–Revised in a longitudinal sample of infant siblings of autistic children (HR; $n = 427$, 171 female, 83.4% White) and a comparison group of low-risk controls (LR, $n = 200$, 86 female, 81.5% White), collected between 2007 and 2017, this study identified an invariant factor structure of temperament traits across groups at 6 and 12 months. Second, after partitioning the groups by familial risk and diagnostic outcome at 24 months, results reveal an endophenotypic pattern of Positive Emotionality at both 6 and 12 months, (HR-autism spectrum disorder [ASD] < HR-no-ASD < LR). Third, increased 'Duration of Orienting' at 12 months was associated with lower scores on the 24-month developmental outcomes in HR infants. These findings may augment efforts for early identification of ASD.

Abbreviations: ADI–R, Autism Diagnostic Interview–Revised; ASD, autism spectrum disorder; CFA, confirmatory factor analysis; CFI, comparative fit index; EFA, exploratory factor analysis; ELC, early learning composite; IBQ–R, Infant Behavior Questionnaire–Revised; MANOVA, multivariate analysis of variance; MSEL, Mullen Scales of Early Learning; NVDQ, nonverbal developmental quotient; RMSEA, root mean square error of approximation; SCQ, Social Communication Questionnaire; SEM, structural equation model; VABS, Vineland Adaptive Behavior Scale; VDQ, verbal developmental quotient.

The collaborators of the IBIS Network are listed in Acknowledgements.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Child Development* published by Wiley Periodicals LLC on behalf of Society for Research in Child Development

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by impairments in social communication and patterns of restricted and repetitive behaviors and interests (American Psychiatric Association, 2013). Given the high heritability associated with ASD, younger siblings of children with the disorder are at a significantly elevated likelihood for receiving a diagnosis relative to the general population (Constantino et al., 2010; Ozonoff et al., 2011). While the reliability and stability of ASD diagnosis prior to 18 months is essentially unknown (Chawarska et al., 2007; Guthrie et al., 2013), a growing body of evidence suggests that early risk markers of ASD are observable by the end of the first year of life (Miller et al., 2017; Ozonoff et al., 2010; Stallworthy et al., 2021; Zwaigenbaum et al., 2021). Thus, prospectively studying infant siblings of autistic children is a valuable method to enhance our understanding of early emerging features of, and developmental trajectories associated with, ASD.

In attempts to identify markers of risk prior to the consolidation of ASD symptoms, temperament has been recognized as a potentially useful construct. Temperament can be defined as early emerging individual differences in activity, reactivity, and regulation that have a biological basis and are shaped by complex interactions between genetic, environmental, and maturational factors (Putnam, 2015; Shiner et al., 2012). Given the strong constitutional foundation of temperament, with “constitutional” referring to relatively enduring biological predispositions, it has been suggested that associations between temperament and developmental outcomes are primarily due to common genetic and environmental factors (Lemery-Chalfant et al., 2008). Therefore, analyses of temperament traits as possible endophenotypes (Gottesman & Gould, 2003), or features that may represent better targets for genetic risk than diagnostic categories, have illustrative potential for disorders with complex inheritance such as ASD (Garon et al., 2016; Putnam, 2015; Shiner et al., 2012).

The broader field of temperament includes several competing models and consequently, varied methods to quantify temperament traits have been developed with a focus on differential characterization of specific temperament models (De Pauw, 2017; Shiner & DeYoung, 2013). Among those, in this study, we focus on Rothbart’s (1981, 2012) model, which is perhaps the most broad and inclusive model that emphasizes both biological reactivity and self-regulation (Shiner & DeYoung, 2013). Developing their questionnaire measures of temperament, Rothbart and colleagues considered other temperament models, as well as adult temperament and personality models to assess a wide range of individual differences (De Pauw & Mervielde, 2010; Shiner & DeYoung, 2013). Reflecting this inclusive strategy, this approach generated detailed measures that have made important contributions in the examination of the higher-order factor structure of temperament. Factor analysis of Rothbart

and colleagues’ questionnaire instruments yielded three higher-order traits, labeled as Surgency/Positive Emotionality, Negative Affect, and Effortful Control/Regulatory Capacity. Surgency captures children’s level of activity, engagement with the environment, and expression of positive emotions. Negative Affect captures tendencies toward negative emotions such as fear, sadness, and anger. And Effortful Control captures individual differences in focusing and sustaining attention and inhibiting dominant responses in favor of a subdominant response (De Pauw, 2017). There is, in general, consensus from conceptual analyses synthesizing the multitude of proposed temperament models that these three overarching traits in part explain the structure of temperament. As Rothbart’s model is among the most widely used in the infancy field, the majority of studies investigating temperament differences between infant siblings of autistic children (those at increased or heightened risk for subsequent disability or impairment associated with receiving a diagnosis or HR) and low-risk (or LR) siblings of typically developing children in the first year of life have used the Infant Behavior Questionnaire (IBQ; Rothbart, 1981) or the IBQ-Revised (IBQ-R; Gartstein & Rothbart, 2003).

Table 1 summarizes previous findings on temperament traits and ASD risk status/diagnosis in infancy. First, HR siblings who were later diagnosed with ASD (HR+) had lower levels of Surgency/Positive Emotionality and related lower-level traits (i.e., Activity Level) compared with HR siblings who did not receive a later diagnosis of ASD (HR-) and LR siblings in most of the studies (Garon et al., 2016; Paterson et al., 2019; Pijl et al., 2019; Zwaigenbaum et al., 2005). There is one discrepant finding where HR+ siblings demonstrated higher level of Surgency than HR- siblings around 7 months (Clifford et al., 2013). However, their follow-up analyses revealed that the difference was observed only in Perceptual Sensitivity among lower-level traits. Second, higher levels of Negative Emotionality and related lower-level traits (i.e., Distress to Limitation and Fear) have been consistently observed in HR+ siblings (Clifford et al., 2013; Garon et al., 2016; Paterson et al., 2019; Pijl et al., 2019; Zwaigenbaum et al., 2005). Lastly, there were mixed findings on Regulatory Capacity and related lower-level traits. HR+ siblings scored lower on Regulatory Capacity as a higher-level trait in most of the studies (Clifford et al., 2013; Paterson et al., 2019; Pijl et al., 2019), but scored higher on a specific lower-level trait, Duration of Orienting (Zwaigenbaum et al., 2005). In sum, on higher-level traits, HR+/HR siblings were rated as having lower levels of Surgency/Positive Emotionality and Regulatory Capacity, but higher levels of Negative Emotionality compared with LR siblings. Some contradictory results in the opposite direction were observed in specific lower-level traits.

Beyond group-mean differences, findings from Garon et al. (2016) suggested that the associations among the

TABLE 1 Summary of findings on temperament traits and autism spectrum disorder risk status/diagnosis in infancy

Study	N	6–10 months			12–14 months			RC
		SU/PE	NE	RC	SU/PE	NE	RC	
Zwaigenbaum et al. (2005)	^a HR (65, 19/46) LR (23)	^a HR+ < HR-, LR				^c HR+ > HR-, LR	^e HR+ > HR-, LR	
Garon et al. (2016)	^a HR (383, 98/285) LR (162)				HR+ < HR-	^f HR > LR		
Clifford et al. (2013)	^b HR (53, 17/12/24) LR (50)	^d HR+ > HR- HR < LR					HR < LR ^h HR+ < HR-, LR	
Pijl et al. (2019)	^b HR (133, 24/34/75) LR (66)		ⁱ HR+ > HR- > LR		HR+ < LR	HR+ > HR- > LR	HR+ < HR- < LR	
Paterson et al. (2019)	^a HR (282, 61/221) LR (114)	HR+ < LR	HR+ > LR	HR+ < HR-, LR	HR+ < HR- < LR	HR+ > LR	HR+ < LR	

Abbreviations: NE, negative emotionality; RC regulatory capacity; SU/PE, surgency/positive emotionality.

^aHR (total, HR+/HR-).

^bHR (total, HR+/HR- with other developmental concerns/HR-).

^cActivity level.

^dPerceptual sensitivity.

^eDistress to limitation.

^fDistress to limitation and fear.

^gDuration of orienting.

^hCuddliness.

ⁱHR+ statistically equivalent to HR-atypical, HR-atypical statistically equivalent to HR-typical.

temperament factors and the predictive relations of the temperament factors to developmental outcomes might differ between the HR and LR groups. Measured by the Toddler Behavior Assessment Questionnaire-Revised (Goldsmith, 1996; Rothbart et al., 2003), the association between Negative Affect and Effortful Control was significant only for the LR group whereas the association between Positive Affect and Effortful Control was significant only for the HR group at 24 months of age. Furthermore, Effortful Control at 24 months predicted ASD symptoms at 36 months only for the HR group.

Although all of the aforementioned studies individually demonstrated interesting associations between temperament and ASD risk status/diagnosis, taken together, findings about temperament in the first year seem mixed or discrepant (Paterson et al., 2019; Pijl et al., 2019). It may be due to analyses conducted on different levels of traits. For example, Zwaigenbaum et al. (2005) found that HR+ siblings had higher scores on Duration of Orienting, a lower-level trait, but other studies (Clifford et al., 2013; Paterson et al., 2019; Pijl et al., 2019) found that HR+ siblings had lower scores on Regulatory Capacity, a higher-level trait including Duration of Orienting. Some inconsistent findings could be attributed, at least in part, to small sample sizes in some of the studies (e.g., Clifford et al., 2013). More importantly, it may reflect differences in the applied constructs across studies. Most of the aforementioned studies examined temperament differences on higher-level traits. However, how those traits were constructed was explicit only in two studies. Paterson et al. (2019) summed *z*-transformed subscale scores to compute factor scores, whereas Garon et al. (2016) derived factor scores from confirmatory factor analysis (CFA) models. Given different methods to compute factor scores in these two studies, traits with the same label might not be readily comparable. Indeed, Negative Affect in Garon et al. (2016) consisted of Activity, Distress to Limitation, and Fear, whereas Negative Affect in Paterson et al. (2019) consisted of Sadness, Distress to Limitation, Fear, and Falling Reactivity.

This question about the comparability of constructs across studies leads to a more fundamental psychometric question of whether temperament constructs, particularly the higher-level traits, measured by the IBQ-R are invariant across HR and LR siblings. To appropriately compare scores across groups, measurement invariance should be examined to ensure that measures reflect the same underlying construct and the construct has the same meaning and structure across groups (Meehl, 1990; Putnick & Bornstein, 2016). Without measurement invariance ensured, group comparisons are invalid, akin to comparing apples with oranges. For example, if Negative Affect mostly reflects Fear and Sadness in LR siblings while it mostly reflects Distress to Limitation and Falling Reactivity in HR siblings, comparing “Negative Affect” scores between two groups would be

theoretically invalid. If measurement invariance does not hold, it also implies that individuals with the same latent-factor scores could have different probabilities of endorsing a response on the measure depending on their group membership (Joo & Kim, 2019). Therefore, measurement invariance is a fundamental question, both empirically and theoretically, in any studies involving group comparisons based on the assumption that the latent structure underlying the psychometric test scores was valid across groups.

Despite the importance of establishing measurement invariance prior to group comparisons in any psychometric test scores, to our knowledge, Garon et al.'s (2016) study is the only one that examined the measurement invariance of temperament constructs and factor structures between HR and LR siblings using infant temperament questionnaires based on Rothbart's model. Considering that Garon et al.'s (2016) study used the IBQ, there is no study that has examined the measurement invariance of the IBQ-R between HR and LR siblings. Thus, in this study, we primarily aim to examine the measurement invariance of the IBQ-R between HR and LR siblings to ensure valid comparisons of group differences in temperament traits. Based on the results of measurement invariance test, we further aim to compare the scores of higher-order traits between HR and LR siblings and to examine whether there exist differential relations among temperament traits, and differential relations of temperament traits to other developmental outcomes.

The current study

The current study consists of three parts. In Part 1, we investigated the invariance of temperament structure between HR and LR siblings using the IBQ-R. In addition to the invariance across groups, we examined the measurement invariance across time as well given that temperament was measured at two time points (6 and 12 months). Because the meaning and structure of temperament traits can change over time, examination of the measurement invariance across time is imperative from a developmental perspective. The results of Part 1 were reflected in subsequent analyses to ensure valid comparisons across groups and time. In Part 2, we examined mean-level differences in higher-order temperament traits by risk status as well as diagnostic outcome at two time points. In Part 3, we investigated (1) differences in the associations among higher-order temperament traits and (2) the extent to which higher-order temperament traits measured in the first year of life (6 and 12 months) differently predict a dimensional developmental outcome related to ASD at 24 months across HR and LR groups using an SEM framework. As the catalyst for this study was a focus on measurement invariance, we consider this study more exploratory than confirmatory in nature.

GENERAL METHODS

Participants

The total sample included 627 (257 females, 370 males) infants; 82.8% were White, 2.7% Black or African American, 1.1% Asian, and 10.7% identified with more than one race/ethnicity. Of these, 427 (171 females, 256 males) were HR infants with an older autistic sibling and 200 (86 females, 114 males) were LR control infants. ASD in older siblings of the HR infants was determined using the Social Communication Questionnaire (SCQ; Rutter et al., 2003) and the Autism Diagnostic Interview–Revised (ADI–R; Lord et al., 1994) prior to the infant enrolling in the study. LR infants who had typically developing older siblings, also screened with the SCQ, were recruited from many sources including: flyers in the community, advertisement in parenting magazines, and outreach at parent events. LR infants were excluded for a family history of a first- or second-degree relative with ASD and/or intellectual disability. Exclusion criteria for both the HR and LR infants included the following: (1) history of known genetic syndromes associated with ASD; (2) significant medical conditions affecting growth, development or cognition or sensory impairments such as significant vision or hearing loss; (3) birth weight <2000 g and/or gestational age <36 weeks; (4) history of significant perinatal adversity, or exposure in-utero to neurotoxins; (5) contraindication for magnetic resonance imaging; (6) predominant home language other than English; (7) having been adopted; and (8) family history of a first-degree relative with psychosis, schizophrenia, or bipolar disorder. The vast majority of infants enrolled at 6 months (a minority enrolled at 12 months) and were assessed in-person at 6 months ($M_{\text{age}} = 6.60$, $SD = 0.96$), 12 months ($M_{\text{age}} = 12.45$, $SD = 3.42$), and 24 months ($M_{\text{age}} = 24.73$, $SD = 1.05$). At the 24-month visit, diagnostic classification of ASD for participants was determined by expert clinical judgment using *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text revised; DSM-IV-TR) criteria and supported by available assessment data including the Autism Diagnostic Observation Schedule (Lord et al., 2000), ADI–R, Vineland Adaptive Behavior Scales (VABS), and the Mullen Scales of Early Learning. Of the HR infants, 81 infants went on to receive an autism diagnosis at 24 months of age and are designated as HR-positive (HR+); HR infants who did not receive a diagnosis at 24 months are designated HR-negative (HR–).

As there was some variability in 6 versus 12-month enrollment, the sample sizes across the three parts of this study vary to some degree. See Tables S1–S4 for sample characteristics and demographic information by risk/diagnostic group, and Tables S5 and S6 for comparisons of children who did and did not contribute data to the analysis in Part 2 and 3.

Measures

The IBQ–R assesses temperament in infants between 3 and 12 months of age. The questionnaire is comprised of 14 subscales with 191 items. Items are scored on a 7-point Likert-type scale ranging from “never” to “always” and are averaged to calculate each subscale score. From these 14 subscales, a three-factor structure is constructed. Surgency/Positive Emotionality factor includes the subscales: Approach, Vocal Reactivity, High-Intensity Pleasure, Smiling and Laughter, Activity Level, and Perceptual Sensitivity. Negative Affect consists of the following subscales: Sadness, Distress to Limitation, Fear, and Falling Reactivity. Effortful Control/Regulatory Capacity encompasses the remaining subscales: Low Intensity Pleasure, Cuddliness, Duration of Orienting, and Soothability. Participants were assessed with the IBQ–R at 6 and/or 12 months of age. Descriptive statistics and Cronbach's alpha for each subscale are shown in Table 2.

The Mullen Scales of Early Learning (MSEL) provides a comprehensive assessment of language, motor, and perceptual abilities for children of all ability levels, ages birth to 68 months (Mullen, 1995). The MSEL provides subscales to assess five developmental areas: (a) gross motor, (b) fine motor, (c) visual reception, (d) expressive language, and (e) receptive language. Early learning composite (ELC) scores are generated from age-normed t-scores of four subscales, fine motor, visual reception, expressive language and receptive language. In addition to the ELC, verbal and nonverbal developmental quotients (VDQ and NVDQ, respectively) were used. VDQ includes scales of expressive language and receptive language and the NVDQ includes scales of visual reception and fine motor. Participants were assessed with the MSEL at 6, 12, and 24 months of age.

The VABS-II assesses child adaptive behavior in the Communication, Socialization, Daily Living Skills, and Motor domains. This measure is commonly used in studies of autism and developmental disabilities and clinical settings to establish an individual's degree of functional impairment. The VABS-II was collected at 24 months of age through a semi-structured interview administered to a parent.

PART 1: THE IBQ–R FACTOR STRUCTURE IN HIGH- AND LOW-RISK INFANTS

Part 1: Method

Participants

Of the total sample, IBQ–R data were available for 518 infants (332 HR, 186 LR) at 6 months and 497 infants (358 HR, 139 LR) at 12 months, with 61.9% contributing data at both time points.

TABLE 2 Means and standard deviations of all scales and tests of group differences

Scales	α	High-risk	Low-risk	Group comparisons			
		<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>F</i>	η^2 [90% CI]	<i>p</i>	Adjusted <i>p</i>
IBQ 6 months							
Activity level	.81	4.35 (0.91)	4.54 (0.84)	1.51	.005 [.00; .02]	.220	.324
Distress to limitation	.81	3.82 (0.84)	3.69 (0.75)	2.03	.006 [.00; .03]	.155	.286
Fear	.90	2.35 (0.89)	2.38 (0.93)	0.16	.000 [.00; .01]	.691	.756
Duration of orienting	.83	3.30 (1.07)	3.19 (1.02)	1.91	.006 [.00; .03]	.168	.294
Smiling and laughter	.83	4.76 (1.06)	5.02 (0.93)	2.30	.007 [.00; .03]	.130	.253
High-intensity pleasure	.82	5.95 (0.74)	6.07 (0.60)	2.54	.008 [.00; .03]	.112	.243
Low-intensity pleasure	.83	5.25 (0.82)	5.37 (0.77)	1.75	.005 [.00; .03]	.186	.310
Soothability	.81	5.09 (0.75)	5.20 (0.70)	0.85	.003 [.00; .02]	.358	.464
Falling reactivity	.84	5.16 (0.88)	5.30 (0.72)	1.50	.005 [.00; .02]	.222	.324
Cuddliness	.89	5.88 (0.73)	5.96 (0.58)	0.00	.000 [.00; 1.00]	.951	.968
Perceptual sensitivity	.84	3.53 (1.20)	3.75 (1.15)	1.51	.005 [.00; .02]	.221	.324
Sadness	.80	3.67 (0.85)	3.50 (0.87)	3.41	.011 [.00; .04]	.066	.196
Approach	.85	5.28 (0.92)	5.34 (0.86)	0.00	.000 [.00; 1.00]	.968	.968
Vocal reactivity	.82	4.31 (1.00)	4.61 (0.85)	3.49	.011 [.00; .04]	.063	.196
ELC 6 months		97.31 (11.66)	100.79 (11.33)	5.18	.016 [.00; .05]	.024	.103
IBQ 12 months							
Activity level	.78	4.36 (0.81)	4.59 (0.84)	3.37	.010 [.00; .04]	.067	.196
Distress to limitation	.82	4.30 (0.84)	4.12 (0.86)	0.46	.001 [.00; .02]	.496	.599
Fear	.89	3.01 (0.96)	2.96 (0.89)	0.30	.001 [.00; .01]	.586	.681
Duration of orienting	.84	2.84 (1.02)	2.95(0.99)	0.02	.000 [.00; .00]	.897	.952
Smiling and laughter	.84	4.86 (1.00)	5.18 (0.85)	2.76	.009 [.00; .03]	.098	.243
High-intensity pleasure	.84	6.09 (0.64)	6.21 (0.53)	1.12	.003 [.00; .02]	.291	.392
Low-intensity pleasure	.82	4.77 (0.89)	4.95 (0.87)	0.74	.002 [.00; .02]	.389	.487
Soothability	.81	5.13 (0.71)	5.31 (0.66)	2.60	.008 [.00; .03]	.108	.243
Falling reactivity	.83	5.21 (0.82)	5.47 (0.67)	5.79	.018 [.00; .05]	.017	.083
Cuddliness	.89	5.39 (0.82)	5.54 (0.64)	1.32	.004 [.00; .02]	.251	.352
Perceptual sensitivity	.85	3.43 (1.14)	3.80 (1.07)	4.66	.014 [.00; .04]	.032	.123
Sadness	.81	3.74 (0.87)	3.51 (0.91)	3.03	.009 [.00; .03]	.083	.223
Approach	.85	5.73 (0.72)	5.83 (0.69)	0.27	.001 [.00; .01]	.603	.681
Vocal reactivity	.84	4.88 (0.92)	5.24 (0.79)	6.92	.021 [.00; .05]	.009	.052
ELC 12 months		99.51 (13.17)	105.04 (12.69)	14.17	.042 [.01; .08]	.000	.002
Vineland 24 months							
Communication		98.83 (12.47)	104.83 (8.85)	19.87	.058 [.02; .10]	.000	.000
Socialization		98.33 (12.28)	102.51 (1.30)	11.18	.034 [.01; .07]	.001	.006
Motor skills		98.72 (12.52)	101.71 (9.47)	2.46	.008 [.00; .03]	.118	.243
Mullen 24 months							
Verbal		94.85(22.33)	107.30 (18.24)	31.25	.089 [.04; .14]	.000	.000
Non-verbal		99.06 (13.96)	107.42 (15.31)	35.52	.100 [.05; .15]	.000	.000

Statistical approach

Assumptions of multivariate normality were evaluated using the Mardia test (Mardia, 1970), which indicated violations of multivariate normality. Therefore, maximum likelihood estimation with robust standard error was used, conducting CFA. Two possible multivariate

outliers were detected using the Mahalanobis distance and Q-Q plot. However, they were not excluded from the analyses because the results yielded without them did not significantly differ from the results with them.

We used three different models to examine the measurement invariance of the IBQ-R between HR and LR siblings (Figure S1): (1) the original three-factor model

suggested by the authors of the IBQ-R (Model 1), (2) a model adopted from the literature (Model 2), and (3) a model derived from our own sample (Model 3). Two alternative models in addition to the original model were examined because the majority of previous studies investigating the structure of infant temperament using the IBQ-R in culturally and sociodemographically diverse samples (Bosquet Enlow et al., 2016; Dragan et al., 2011; Gartstein et al., 2005) have used modified models given the poor fit of the original model in diverse samples. Therefore, we adopted the second model that was proposed by Gartstein et al. (2005) using a pooled US sample ($N = 608$), which excluded Cuddliness (due to low factor loadings) and High-Intensity Pleasure (due to substantial residual covariation), and was relaxed allowing for several secondary loadings and error covariance between items.

The third model was derived from our total sample (LR and HR collapsed) using a split-sample approach. To ensure comparability of analytic subsamples, the sample was first stratified by sex and risk status and randomly assigned to one of two groups: the exploratory factor analysis (EFA) sample ($N = 314$) and the CFA sample ($N = 313$). First, EFA was performed enforcing a three-factor solution and CIs for factor loadings were constructed using the nonparametric bootstrapping. Factor cross-loadings and residual covariance were further examined using modification indices adopting EFA within CFA framework. In this EFA model, for a clearer interpretation of each factor, we (1) excluded scales that did not clearly load onto any factors and scales that did not load onto the supposed factor, but did load onto other factors, (2) allowed only one secondary loading, and (3) allowed residual covariance only among indicators within the same factor. Subsequently, we ran CFA testing the fit of the model generated from our EFA. See Figure S2 for detailed results of EFA and CFA. Finally, the full sample was used to assess measurement invariances of all three models.

Before proceeding to the measurement invariance test, whether each model fits the data well for HR and LR respectively was tested (single-group test). The model fit was evaluated by four indices: the χ^2 test, the comparative fit index (CFI), and the root mean square error of approximation (RMSEA). Values greater than .90 for CFI, and values less than .08 for RMSEA were considered acceptable fit (Little, 2013). All confirmatory factor analyses were performed using the R-package lavaan (Rosseel, 2012). The results indicated that Model 3 had a good fit to the data for both HR and LR groups at both time points, Model 2 had a good fit to the data except for HR at 12 months of age, and Model 1 had a poor fit for the both groups at all time points. More detailed information on the fit indices of single-group tests for all three models can be found in Table S7.

The measurement invariance across the HR and LR groups was tested in a CFA framework, comparing fit

to the data of a hierarchy of models with increasingly stringent constraints imposed on the factor structure. Our analysis tested the following four steps of group invariance: (1) configural invariance, which requires the same patterns of general factor structure, (2) metric invariance, which requires the same factor loading matrices across groups, (3) scalar invariance, which constrains both factor loadings and intercepts of indicators to be the same across groups, and (4) residual invariance, which further constrains the measurement error covariance matrices of indicators to be the same across groups (Widaman & Reise, 1997).

Part 1: Results

Group invariance of infant temperament factor structure

Table 3 provides a summary of the results of the group measurement invariance testing. Model 1 demonstrated poor fit even when no equality constraints were applied, suggesting that the pattern of fixed and freed factor loadings was not similar across the groups at both time points. Given no support for configural invariance, no further interpretations on measurement invariance were made.

Model 2, proposed by Gartstein et al. (2005), demonstrated adequate fit when unconstrained loadings were estimated, providing support for configural invariance at both time points. Constraining the factor loadings to be equivalent yielded adequate fit and the difference between the models was not significant, suggesting that factor loadings were equivalent for HR and LR groups at both time points. Furthermore, constraining the intercepts and error variance to be equivalent across the groups also produced adequate fit and statistically non-significant model comparison results at 6 months, suggesting that HR and LR groups did not differ on either subscale intercept or error variance at 6 months. At 12 months, however, a scalar invariance test produced a significantly poorer fit, suggesting that the groups differ on subscale intercepts.

Model 3, derived from our sample, demonstrated adequate fit for configural invariance, providing support for the same general factor structure across the groups at both time points. Further tests of metric invariance and scalar invariance revealed that factor loadings and subscale intercepts were equivalent across the groups at both time points. Error variance, on the other hand, differed between the groups at 6 months.

Although both Model 2 and Model 3 satisfied measurement invariance up to the level of scalar invariance, we adopted Model 3 for the further analyses given its simpler structure, which would be more generalizable. Figure 1 shows Model 3 at each time point with all coefficients freely estimated.

TABLE 3 Summary of measurement invariance tests

	<i>df</i>	χ^2	$\Delta\chi^2$	<i>p</i> ^a	CFI	RMSEA	AIC	BIC
Model 1								
6-month								
Configural invariance	148	625.10			.765	.112	17,074.50	17,457.00
Metric invariance	162	648.12	23.02	.059	.760	.108	17,069.52	17,392.52
Scalar invariance	173	663.03	14.91	.187	.758	.105	17,062.43	17,338.68
Residual invariance	187	688.65	25.62	.028	.753	.102	17,060.05	17,276.80
12-month								
Configural invariance	148	709.33			.730	.124	15,999.07	16,377.84
Metric invariance	162	740.68	31.34	.005	.722	.120	16,002.41	16,322.27
Scalar invariance	173	761.29	20.61	.038	.717	.117	16,001.02	16,274.58
Residual invariance	187	787.33	26.04	.026	.712	.114	15,999.06	16,213.70
Model 2								
6-month								
Configural invariance	84	201.70			.922	.074	15,088.61	15,496.61
Metric invariance	103	224.59	22.89	.242	.920	.068	15,073.50	15,400.75
Scalar invariance	112	239.00	14.42	.108	.916	.066	15,069.91	15,358.91
Residual invariance	124	255.16	16.16	.184	.913	.064	15,062.07	15,300.07
12-month								
Configural invariance	84	238.22			.901	.086	14,042.04	14,446.07
Metric invariance	103	259.45	21.23	.324	.900	.078	14,025.27	14,349.33
Scalar invariance	112	285.29	25.84	.002	.889	.079	14,033.11	14,319.30
Residual invariance	124	306.15	20.86	.052	.883	.077	14,029.97	14,265.65
Model 3								
6-month								
Configural invariance	42	118.33			.938	.084	10,972.71	11,253.20
Metric invariance	52	132.61	14.28	.161	.934	.077	10,966.99	11,204.99
Scalar invariance	58	142.97	10.36	.110	.931	.075	10,965.35	11,177.85
Residual invariance	66	160.76	17.79	.023	.923	.074	10,967.14	11,145.64
12-month								
Configural invariance	42	157.93			.914	.105	9996.61	10,274.37
Metric invariance	52	174.95	17.02	.074	.909	.098	9993.63	10,229.31
Scalar invariance	58	182.38	7.44	.282	.908	.093	9989.06	10,199.49
Residual invariance	66	196.65	14.27	.075	.903	.089	9987.34	10,164.10

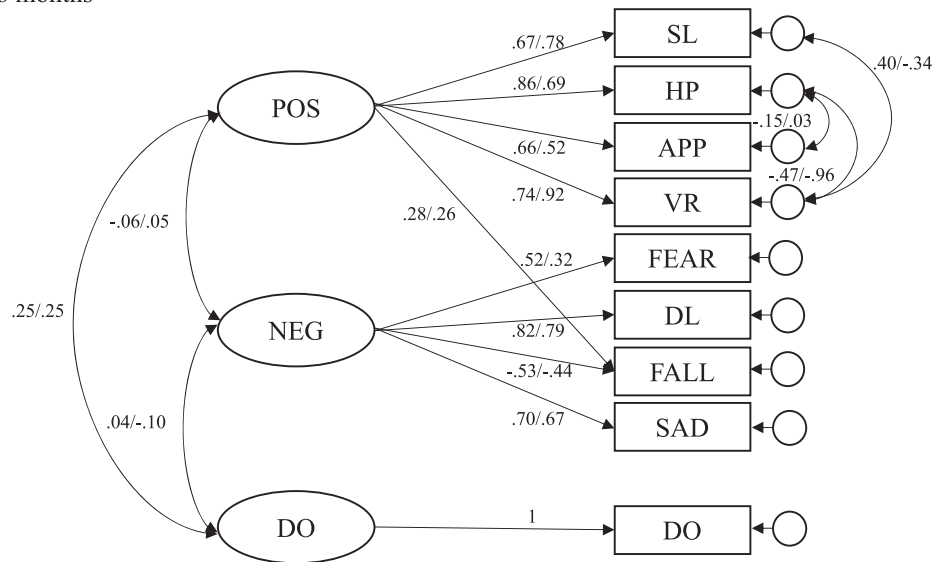
^aChi-square difference test.

Longitudinal invariance in the structure of infant temperament

As the data involves longitudinal temperament assessments at 6 and 12 months, we also tested measurement invariance across time in addition to groups. A longitudinal model, which included associations between corresponding factors and subscales across two time points without any constraints on loadings or intercepts, demonstrated adequate fit, $\chi^2(216) = 491.976$, $p < .001$, CFI = .917, RMSEA = .065. Constraining the factor loadings to be equivalent across time and

groups yielded adequate fit, $\chi^2(257) = 538.711$, $p < .001$, CFI = .917, RMSEA = .060, and the difference between the models was not significant, $\Delta\chi^2(41) = 46.735$, $p = .249$, Δ CFI = .001. Next, scalar invariance was tested. Constraining intercepts of corresponding subscales to be equal between 6- and 12-months and groups yielded poor fit (Little, 2013), $\chi^2(263) = 868.250$, $p < .001$, CFI = .822, RMSEA = .087, and the difference from the metric invariance model was statistically significant, $\Delta\chi^2(6) = 329.539$, $p < .000$, Δ CFI = .095. As scalar invariance was not satisfied across two time points, comparisons of factor means were made only across groups

(a) 6 months



(b) 12 months

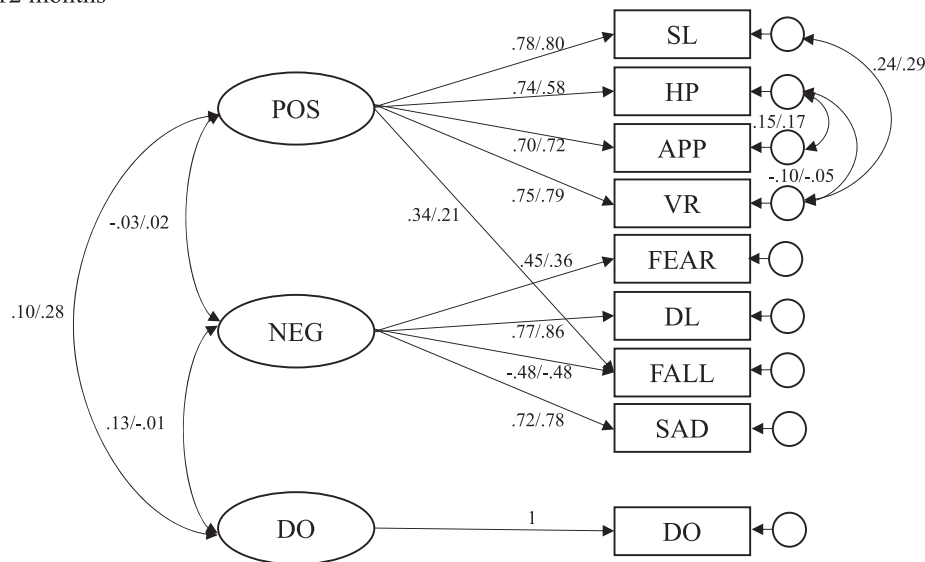


FIGURE 1 Exploratory factor analysis (EFA) models of the IBQ-R structure in the current sample. Standardized estimates are presented. High-risk/low-risk groups. POS, Positive Emotionality; NEG, Negative Emotionality; DO, Duration of Orienting; SL, Smiling and Laughter; HP, High-Intensity Pleasure; APP, Approach; VR, Vocal Reactivity; DL, Distress to Limitations; FALL, Falling Reactivity; SAD, Sadness

within a time point in Part 2, and metric invariant model was used for a longitudinal model in Part 3.

PART 2: TEMPERAMENT DIFFERENCES BY DIAGNOSTIC STATUS

Part 2: Method

Participants

Diagnostic outcomes based on DSM-IV-TR criteria at 24 months were available for 507 participants (96.4% of the sample from Part 1). Based on risk and diagnostic status,

three groups were identified: low-risk, not diagnosed with ASD (LR-; $N = 146$), high-risk, not diagnosed with ASD (HR-; $N = 280$), and high-risk, diagnosed with ASD (HR+; $N = 81$). Low-risk children subsequently diagnosed with ASD ($n = 3$) were excluded to maintain the family design.

Statistical approach

Temperament factor scores were computed using Model 3 from Part 1, constraining factor loadings and intercepts of corresponding subscales to be equal across groups within each time point. A multivariate analysis of variance (MANOVA) was used with group as

the independent variables, and temperament factors (Positive Emotionality, Negative Emotionality, and Duration of Orientating) as the dependent variables.

Part 2: Results

Figure 2 shows the temperament factor scores for the three groups. Results of the MANOVA indicated a significant group effect for the temperament factor scores, $F(2504) = 5.66$, $h^2 = .06$, 90% CI [.00; .05], $p < .001$. After correction for multiple comparisons using the Benjamini–Hochberg procedure, results of F -tests for each temperament factor score indicated that Positive Emotionality and Negative Emotionality at 6 and 12 months differed by group. The HR+ group had significantly lower scores on Positive Emotionality compared with both the LR– group and HR– group; $t(225) = 6.04$, $p < .001$, $d = .82$, 95% CI [.55; 1.09] and $t(359) = 3.73$, $p < .001$, $d = .47$, 95% CI [.26; .68] at 6 months and $t(225) = 7.26$, $p < .001$, $d = 1.00$, 95% CI [.72; 1.28] and $t(359) = 4.71$, $p < .001$, $d = .59$, 95% CI [.40; .82] at 12 months. When comparing the LR– group and HR– group, the HR– group had significantly lower scores on Positive Emotionality at both 6 and 12 months, $t(424) = 3.02$, $p < .01$, $d = .32$, 95% CI [.12; .51] and $t(424) = 3.19$, $p < .01$, $d = .33$, 95% CI [.14; .52], respectively. The HR+ group had significantly higher scores on Negative Emotionality compared with LR– group; $t(225) = 2.89$, $p < .05$, $d = .41$, 95% CI [.15; .67] at 6 months and $t(225) = 2.97$, $p < .01$, $d = .43$, 95% CI [.17; .69] at 12 months. The HR+ group, however, did not differ from the HR– group in Negative Emotionality. The LR– group had significantly lower scores on Negative

Emotionality compared with the HR– group only at 6 months, $t(424) = 2.45$, $p < .05$, $d = .25$, 95% CI [.06; .45].

PART 3: LONGITUDINAL MODEL PREDICTING DEVELOPMENTAL OUTCOMES AT 24 MONTHS

Part 3: Method

Participants

The sample included 526 infants (214 females, 312 males) who completed the MSEL and/or VABS-II at 24 months. The majority of the sample (91.3%) completed both assessments. Of these, 368 were HR infants and 158 were LR control infants.

Statistical approach

Given that the assessments had different ranges, each variable was re-scaled to range from 0 to 1, using the formula: $(x_i - \min(x))/(\max(x) - \min(x))$. Preliminary analyses indicated that metric invariance held across males and females, fully at 6 months and partially at 12 months with 3 loadings freely estimated. Therefore, sex was included in the longitudinal model only as a covariate.

For developmental outcomes at 24 months, three subscales of the VABS-II (Communication, Socialization, and Motor) and two scores derived from the MSEL (VDQ and NVDQ) were used to generate a data-driven latent outcome variable. A one-factor model with factor

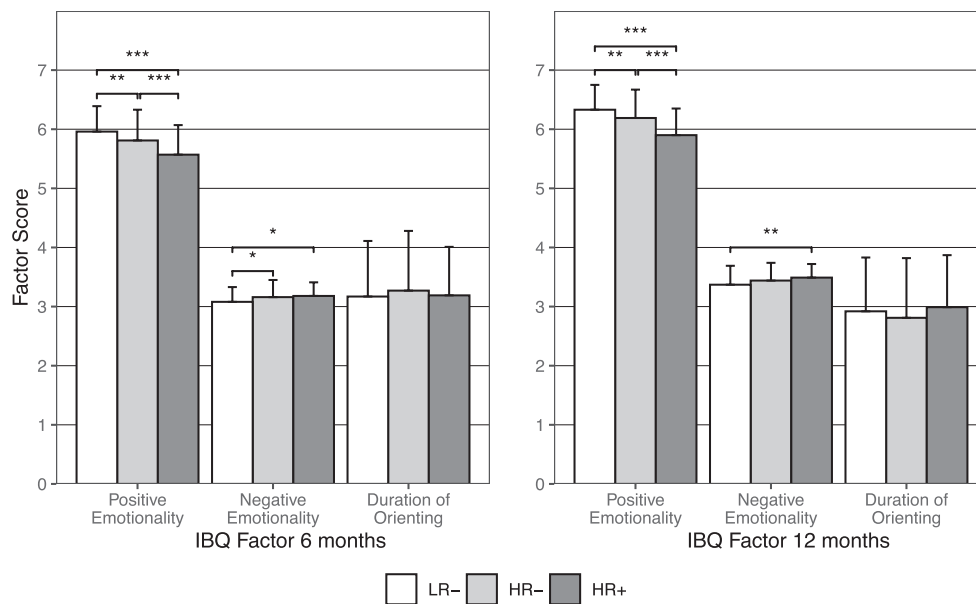


FIGURE 2 Factor scores by risk status and diagnostic status. *Note:* Scores on the temperament factors and Duration of Orienting as a function of familial risk and diagnostic status at 24 months. Asterisks represent p values $< .05$, $.01$, and $.001$, respectively. Error bars represent 1 SD . IBQ, Infant Behavior Questionnaire

loadings constrained to be equal across the HR and LR groups demonstrated good fit, $\chi^2(15) = 74.96$, $p < .000$, CFI = .919, RMSEA = .147, providing evidence that the factor loadings across the groups were invariant.

Ideally, diagnostic outcome would also be incorporated into this model. However, the HR+ sample size was insufficient for inclusion as an independent group. In an attempt to determine whether the differences observed between the HR and LR groups in this SEM may have been driven primarily by the presence of children with ASD, a second SEM was constructed which excluded the HR+ infants. All analyses were performed using the R-package lavaan (Rosseel, 2012) and maximum likelihood estimation with robust standard error was used.

Part 3: Results

Predicting developmental outcomes at 24 months

Using a structural equation model (SEM), we examined the association between 6- and 12-month temperament factors, sex, ELC at 6 and 12 months, and the developmental outcome factor (described above) in a longitudinal model incorporating invariance constraints on factor loadings across time and group. The full model showed mediocre fit (Little, 2013), $\chi^2(526) = 1103.052$, $p < .001$, CFI = .871, RMSEA = .065. Direct paths from 6-month temperament factors to 24-month developmental outcome factor were not significant for both groups, and hence trimmed from the model. The trimmed model also showed mediocre fit (Little, 2013), $\chi^2(534) = 1105.858$, $p < .001$, CFI = .872, RMSEA = .064, and did not significantly differ from the full model, $\Delta\chi^2(8) = 2.806$, $p = .946$, $\Delta\text{CFI} = -.001$. The trimmed model, in fact, showed better fit using relative fit index (e.g., CFI) as Δdf was greater than $\Delta\chi^2$ between the models. Figure 3 shows the trimmed model and coefficients for both groups.

For both HR and LR groups, 12-month temperament factors were predicted by respective 6-month temperament factors, with all coefficients being significant, $p < .001$, which indicated continuity within the constructs over this time period. With regard to the association among the temperamental factors, Positive Emotionality and Duration of Orienting at 6 months were significantly associated for both HR ($r = .26$, $p < .000$) and LR ($r = .28$, $p = .004$) groups. At 12 months, however, the association between Positive Emotionality and Duration of Orienting was significant only for the LR group, $r = .24$, $p = .047$.

ELC at 12 months was significantly predicted by ELC at 6 months for both HR ($\beta = .22$, $p < .01$) and LR ($\beta = .29$, $p < .001$) groups. At 6 months, there was a significant association between ELC and Positive Emotionality for both HR ($r = .25$, $p = .001$) and LR ($r = .42$, $p = .001$) groups. At 12 months, on the other hand, ELC was significantly

associated with Negative Emotionality only for the HR group, $r = .19$, $p = .009$.

For the HR group, the developmental outcomes at 24 months was significantly predicted by four variables: sex, 12-month ELC, 12-month Positive Emotionality, and 12-month Duration of Orienting. Being female predicted higher developmental outcome scores at 24 months ($\beta = -.17$, $p = .001$). Higher scores on ELC at 12 months predicted higher scores on the developmental outcomes at 24 months ($\beta = .28$, $p < .001$). Interestingly, higher scores on Positive Emotionality predicted higher scores on the developmental outcomes at 24 months ($\beta = .22$, $p = .004$), but lower scores on Duration of Orienting at 12 months predicted higher scores on the developmental outcomes at 24 months ($\beta = -.20$, $p = .002$). For the LR group, sex and 12-month ELC significantly predicted the developmental outcomes at 24 months. Being female predicted higher developmental outcome scores at 24 months ($\beta = -.13$, $p = .040$). Higher scores on ELC at 12 months predicted higher scores on the developmental outcomes at 24 months ($\beta = .31$, $p < .001$).

Finally, we tested the SEM model using 6- and 12-month temperamental profiles excluding HR+ children to test whether the differences observed between the HR and LR groups were driven by HR+ children. Figure S3 shows the model with coefficients for both the HR- and LR groups. Most of the coefficients remained significant in the model excluding HR+ children. Importantly, developmental outcome at 24 months was significantly predicted only by Duration of Orienting at 12 months for the HR- group, $\beta = -.20$, $p = .002$ and the associations between temperament factors at 12 months were still significant, albeit with an alpha level of .10, only in the LR- group ($r = .21$, $p = .087$), which demonstrated that the different associations among the HR and LR groups mostly remained unchanged excluding HR+ children.

DISCUSSION

In the present study, we characterized differences in temperamental profiles and the predictive value of early temperament measures for developmental outcome in infants at high and low risk for developing autism on solid psychometric grounds with measurement invariance ensured.

Part 1

We tested measurement invariance of the IBQ-R and the results revealed that (1) temperament factors derived from the IBQ-R reflected the same underlying constructs and (2) the constructs have the same meaning and structure between HR and LR groups. To test measurement invariance, we compared two data-driven modified models along with the original IBQ-R factor model. The poor fit of the original factor model, as demonstrated in previous studies,

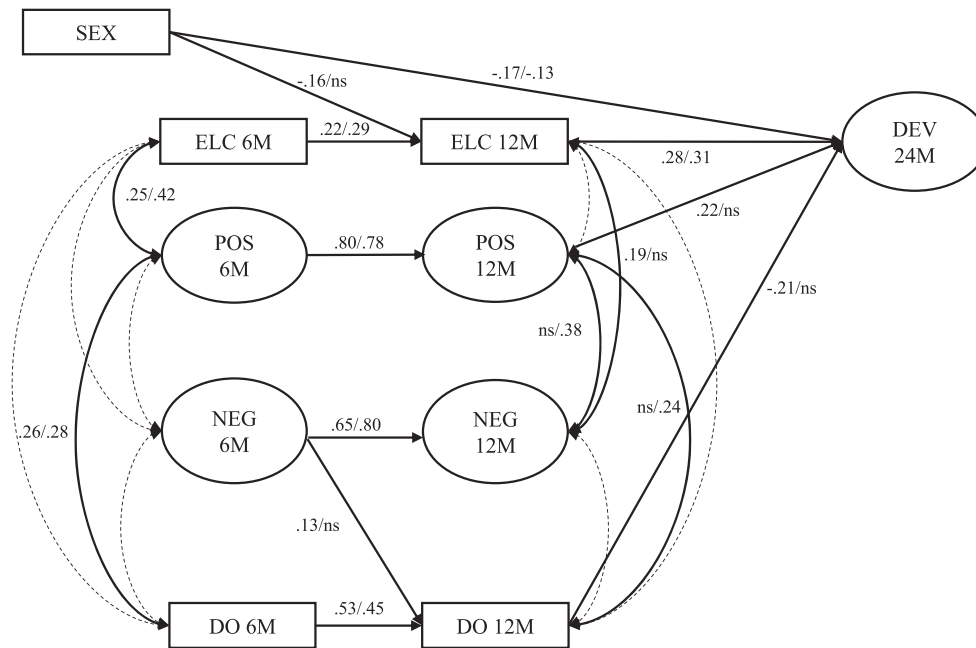


FIGURE 3 The structural equation model predicting 24-month developmental outcome. *Note:* High-risk/low-risk groups. Non-significant direct paths from 6 to 24 months are dropped from the model. Standardized path coefficients that are statistically significant at $p < .05$ are presented except for the path from Sex to Duration of Orienting at 12 months in low-risk group ($b = -.16, p = .048$). Non-significant correlations in both groups are displayed as dotted lines in the figure. DO, Duration of Orienting; ELC, early learning composite; NEG, Negative Emotionality; POS, Positive Emotionality

was replicated in our sample. Two data-driven modified models, however, both satisfied configural, metric, and scalar invariance. This result validated further comparisons of mean scores of the IBQ-R factors and their predictive relations to later developmental outcomes. Although measurement invariance tests showed utility for both models, Model 3, derived from our sample, was selected for subsequent analyses because it had a simpler, hence more interpretable and generalizable, structure with a fewer cross-loadings (1 vs. 5) and residual covariance allowed among indicators of different factors (0 vs. 2). Negative Emotionality in Model 2, particularly, was difficult to interpret and generalize to other studies because four out of five indicators of Surgency/Positive Emotionality—Activity, Smiling and Laughter, Perceptual Sensitivity and Approach—also loaded on Negative Emotionality.

Part 2

Using the scalar invariant model of Model 3, the factor scores of the IBQ-R were compared between groups subdivided into LR-, HR-, and HR+, and a MANOVA indicated a significant group effect on the factor scores. The HR+ group showed lower Positive Emotionality scores than both the HR- and LR groups at both 6 and 12 months, providing evidence of a disorder specific effect. Additionally, the LR group significantly differed from the HR- group at 6 and 12 months, suggestive of an endophenotype pattern of results (LR > HR- > HR+). The LR group showed lower Negative Emotionality

scores than both the HR- and HR+ groups, providing putative evidence of a familial pattern.

This is not the first study to indicate lower positive affect in both high risk infants (Clifford et al., 2013) or infants who go on to receive an autism diagnosis (Garon et al., 2016). In the current study, however, prospective parent report has revealed a disorder-specific difference in temperamental profiles that is evident by 6 months of age, a time prior to the emergence of many of the behavioral signatures associated with autism. Similar to recent work using a novel parent report version of the Autism Observation Scale for Infants (Sacrey et al., 2018), this finding points to the prognostic value of parent report questionnaires in the first year of life. This and other evidence, much of which comes from the high-risk infant sibling literature, implicates the second half of the first year of life as a particularly critical window for understanding the development of autism. There is no evidence of overt observable behavioral features of autism at 6 months of age. A small number of studies reveal overt observable behavioral features at 9 or 10 months of age (Miller et al., 2017; Nyström et al., 2018; Stallworthy et al., 2021). But by 12 months of age, an increased number of overt behavioral symptoms associated with autism are evident, indexed at the group level, such as differences in emerging language/communication abilities (Estes et al., 2015; Ozonoff et al., 2014; Zwaigenbaum et al., 2021) and repetitive behaviors (Elison et al., 2014; Ozonoff et al., 2008; Wolff et al., 2014). Certain temperamental features, such as lower positive affect, may

represent a pre-symptomatic risk marker for autism, but this finding warrants replication.

Part 3

Using the metric invariant model of Model 3, we examined whether the IBQ-R factors differently predicted later developmental outcomes at 24 months of age. Here, we used a composite variable created by combining measures of adaptive and cognitive function to derive a continuous, latent outcome measure. Notably, there were no direct 6-month predictors of 24-month outcome; rather, this relationship was mediated by 12-month development. Temperament factors at 12 months were predicted by 6-month temperament, suggesting that the IBQ-R may indeed be tapping into similar constructs over time, even if those constructs show different patterns of associations with other variables over development.

There were, however, some differences between the LR and HR groups. Both groups showed a significant positive association between ELC scores and Positive Emotionality at 6 months; at 12 months, Negative Emotionality was positively associated with ELC in just the HR group. At 6 months, a significant positive association between Positive Emotionality and Duration of Orienting was observed in both the HR and LR groups; at 12 months, this relationship was significant only for the LR group. It is unsurprising that higher 12-month ELC predicted higher 24-month developmental outcome in the HR group; more interesting is the finding that higher Duration of Orienting predicted lower developmental outcome at 24 months, which occurred only in the HR group. The impetus for re-running the SEM without the HR+ sample was to investigate whether the most compelling difference from the full SEM were driven primarily by the presence of individuals who would go on to receive an autism diagnosis. Indeed, the significant negative relationship between 12-month Duration of Orienting and 24-month outcome did remain. These results suggest that duration of orienting may represent a broader familial risk marker. While this may seem contrary to findings from other studies, which have shown an association between ASD and lower effortful control, it is vital to keep in mind that the model used in this analysis excluded all regulatory subscales aside from Duration of Orienting. The items assessed in the Duration of Orienting subscale (e.g. "How often during the last week did the baby stare at a mobile, crib bumper, or picture for 5 min or longer?") seem to reflect aspects of sustained or perhaps perseverative attention, specifically.

In normative samples, "duration of orienting" appears to decline over the first year of life (Gartstein & Rothbart, 2003). This decrease is consistent with a relative increased engagement of top-down endogenous control attention systems across the latter half of the first

year (Colombo & Cheatham, 2006), more flexible and efficient visual orienting skills, and increases in information processing speed. This shift is typically attributed to maturation in the attentional system including development of a fronto-parietal brain network associated with spatial orienting (Petersen & Posner, 2012; Posner & Rothbart, 1991). Atypical visual orienting in infancy has been implicated in the early developmental course of autism (Bryson et al., 2018; Elison et al., 2013; Elsabbagh et al., 2013; Zwaigenbaum et al., 2005). These abnormalities in orienting among HR+ infants have been associated with abnormal functional specialization of posterior cortical brain circuits (Elison et al., 2013). In this context, the association between higher duration of orienting and poorer developmental outcome observed in the current HR+ sample may reflect delayed or altered specialization of attentional networks.

The relationship between positive affect and duration of orienting may also be an important piece of this story. Adult studies have shown that approach-motivated positive affect reduces the breadth of attention or increases focused attention (Gable & Harmon-Jones, 2008; Harmon-Jones & Gable, 2009), which might help explain the positive association between Positive Emotionality and Duration of Orienting observed in both groups at 6 months of age. However, in our study only the LR group showed an association between Positive Emotionality and Duration of Orienting at 12 months of age. This finding complements a recent report showing the important contribution of shared positive affect in the context of joint attention episodes on subsequent sustained attention (Yu & Smith, 2016). Sustained or focused attention without the correlated support of positive affect, as observed in the HR group at 12 months, may yield different levels of processing/engagement. The question as to when sustained or focused attention becomes perseverative and maladaptive is open for future inquiry.

Limitations

One important limitation of this study is that the HR+ sample size was insufficient to test the SEM using diagnostic outcome, rather than simply high versus low risk for ASD. The SEM in Part 3 was re-run after excluding the sample of HR+ infants (as well as infants for whom the DSM-IV checklist was not completed), and the association between Duration of Orienting and the latent developmental outcome remained the same. This suggests that the SEM findings were not driven by diagnostic outcome, but by risk status; that is, the broader temperamental profiles observed in the HR group may be indicative of familial risk.

A second potential limitation is that diagnostic outcome, in this study, is assessed at 24 months of age. There is no gold standard age of diagnosis for ASD, but there is evidence that a proportion of HR children who do not

meet criteria for a clinical-best-estimate diagnosis at 24 months, will at 36 months of age (Ozonoff et al., 2015; Zwaigenbaum et al., 2016). In addition, these temperamental factors are hypothesized to be relatively stable, but we know less about the stability of the ASD symptom profile in this age range and how this variability relates to dimensions of temperament.

Another important limitation is that this study included primarily White-Non-Hispanic, and upper-middle income class children. Although it is difficult to address in this sample considering the longitudinal nature of the study, including racially, ethnically, and socioeconomically diverse children in a future study will strengthen the generalizability of the findings.

Although it was beyond the scope of this paper, analysis of more fine-grained results, such as individual subscales, would likely present compelling insights (see Paterson et al., 2019). In addition, while the finding relating lower positive affect at 6 months of age to autism diagnostic outcome at 24 months is scientifically compelling, it is less clinically meaningful without individual level prediction. Practically speaking, this suggests that the IBQ-R, or subscales therein could be employed as part of a battery geared towards such prediction. Future work should also incorporate samples enriched for risk for other neurodevelopmental disorders, such as intellectual disabilities, Down syndrome, ADHD, or fragile X syndrome, to characterize the specificity of the temperamental features or profiles associated with autism.

CONCLUSIONS

We established measurement invariance of the IBQ-R factor structure using two data-driven modified models and, meanwhile, provided a model that fit our sample of infants at high and low risk of autism at both 6 and 12 months of age, which will benefit from follow-up studies using CFA in other samples. We also report that 6-month-old high-risk infant siblings subsequently diagnosed with autism show lower Positive Emotionality compared with both high- and low-risk infants not diagnosed with autism at 24 months. An SEM using this factor structure revealed that sex, Positive Emotionality/Surgency, and Duration of Orienting were associated with a dimensional risk outcome in HR siblings at 24 months of age, of which the Duration of Orienting finding appears specific to familial risk. These findings, coupled with recent work from the Canadian Infant Sibling Study (Sacrey et al., 2018) highlight the value of parent report in identifying early signs that differentiate infants who later develop autism. These findings have implications for (1) establishing psychometric evaluation priorities for using the IBQ-R for group comparisons of children enriched for ASD risk, and (2) future efforts aimed at

population-based screening for ASD risk in the first year of life using early temperament traits.

ACKNOWLEDGMENTS

We thank the families who contributed their time and effort to the IBIS study. This work was supported by an NIH Autism Centers of Excellence Network grant R01 HD055741 (Piven), a grant from Autism Speaks #6020 (Piven), a grant from the Simons Foundation #140209 (Piven), and U54 Intellectual and Developmental Disabilities Research Centers (IDDR) at each clinical site (HD079124 to UNC, Piven; HD087011 to Washington University, Constantino & Schlaggar; HD86984 to CHOP, Yudkoff; HD083091 to University of Washington, Guralnick). J. Elison is supported by R01 MH104324 and J. Wolff is supported by R01 MH116961. The Infant Brain Imaging Study (IBIS) Network is an NIH funded Autism Centers of Excellence project and consists of a consortium of 9 universities in the U.S. and Canada. Clinical Sites: University of North Carolina: J. Piven (IBIS Network PI), H. C. Hazlett, C. Chappell; University of Washington: S. Dager, A. Estes, D. Shaw; Washington University: K. Botteron, R. McKinstry, J. Constantino, J. Pruett; Children's Hospital of Philadelphia: R. Schultz, J. Pandey; University of Alberta: L. Zwaigenbaum; University of Minnesota: J. T. Elison, J. Wolff. Data Coordinating Center: Montreal Neurological Institute: A. C. Evans, D. L. Collins, G. B. Pike, V. Fonov, P. Kostopoulos, S. Das, L. MacIntyre. Image Processing Core: New York University: G. Gerig; University of North Carolina: M. Styner. Statistical Analysis Core: University of North Carolina: H. Gu.

DATA AVAILABILITY STATEMENT

The raw data used for this study can be accessed through the NIH National Data Archive. The analysis scripts are available at <https://github.com/sungx077/IBQ-R-factor-structure-autism>.

ORCID

Sooyeon Sung  <https://orcid.org/0000-0001-8557-1168>
 Jason J. Wolff  <https://orcid.org/0000-0003-4300-8195>
 Jed T. Elison  <https://orcid.org/0000-0002-6246-1000>

REFERENCES

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5*. American Psychiatric Association.
- Bosquet Enlow, M., White, M. T., Hails, K., Cabrera, I., & Wright, R. J. (2016). The Infant Behavior Questionnaire-Revised: Factor structure in a culturally and sociodemographically diverse sample in the United States. *Infant Behavior and Development, 43*, 24–35. <https://doi.org/10.1016/j.infbeh.2016.04.001>
- Bryson, S., Garon, N., McMullen, T., Brian, J., Zwaigenbaum, L., Armstrong, V., Roberts, W., Smith, I., & Szatmari, P. (2018). Impaired disengagement of attention and its relationship to emotional distress in infants at high-risk for autism spectrum disorder. *Journal of Clinical and Experimental Neuropsychology, 40*, 487–501. <https://doi.org/10.1080/13803395.2017.1372368>

- Chawarska, K., Klin, A., Paul, R., & Volkmar, F. (2007). Autism spectrum disorder in the second year: Stability and change in syndrome expression. *Journal of Child Psychology and Psychiatry*, *48*, 128–138. <https://doi.org/10.1111/j.1469-7610.2006.01685.x>
- Clifford, S. M., Hudry, K., Elsabbagh, M., Charman, T., & Johnson, M. H. (2013). Temperament in the first 2 years of life in infants at high-risk for autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *43*, 673–686. <https://doi.org/10.1007/s10803-012-1612-y>
- Colombo, J., & Cheatham, C. L. (2006). The emergence and basis of endogenous attention in infancy and early childhood. *Advances in Child Development and Behavior*, *34*, 283–322.
- Constantino, J. N., Zhang, Y., Frazier, T., Abbacchi, A. M., & Law, P. (2010). Sibling recurrence and the genetic epidemiology of autism. *American Journal of Psychiatry*, *167*, 1349–1356. <https://doi.org/10.1176/appi.ajp.2010.09101470>
- De Pauw, S. S. (2017). Childhood personality and temperament. The Oxford handbook of the five factor model. In T. Widiger (Ed.), *The Oxford handbook of the five factor model* (pp. 243–280). Oxford University Press.
- De Pauw, S. S. W., & Mervielde, I. (2010). Temperament, personality and developmental psychopathology: A review based on the conceptual dimensions underlying childhood traits. *Child Psychiatry & Human Development*, *41*, 313–329. <https://doi.org/10.1007/s10578-009-0171-8>
- Dragan, W. L., Kmita, G., & Fronczyk, K. (2011). Psychometric properties of the Polish adaptation of the Infant Behavior Questionnaire—Revised (IBQ-R). *International Journal of Behavioral Development*, *35*, 542–549. <https://doi.org/10.1177/0165025411422181>
- Elison, J. T., Paterson, S. J., Wolff, J. J., Reznick, J. S., Sasson, N. J., Gu, H., Botteron, K. N., Dager, S. R., Estes, A. M., Evans, A. C., Gerig, G., Hazlett, H. C., Schultz, R. T., Styner, M., Zwaigenbaum, L., & Piven, J. (2013). White matter microstructure and atypical visual orienting in 7-month-olds at risk for autism. *The American Journal of Psychiatry*, *170*, 899–908. <https://doi.org/10.1176/appi.ajp.2012.12091150>
- Elison, J. T., Wolff, J. J., Reznick, J. S., Botteron, K. N., Estes, A. M., Gu, H., Hazlett, H. C., Meadows, A. J., Paterson, S. J., Zwaigenbaum, L., & Piven, J. (2014). Repetitive behavior in 12-month-olds later classified with autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *53*, 1216–1224. <https://doi.org/10.1016/j.jaac.2014.08.004>
- Elsabbagh, M., Fernandes, J., Jane Webb, S., Dawson, G., Charman, T., & Johnson, M. H. (2013). Disengagement of visual attention in infancy is associated with emerging autism in toddlerhood. *Biological Psychiatry*, *74*, 189–194. <https://doi.org/10.1016/j.biopsych.2012.11.030>
- Estes, A., Zwaigenbaum, L., Gu, H., St. John, T., Paterson, S., Elison, J. T., Hazlett, H., Botteron, K., Dager, S. R., Schultz, R. T., Kostopoulos, P., Evans, A., Dawson, G., Eliason, J., Alvarez, S., & Piven, J. for the IBIS Network. (2015). Behavioral, cognitive, and adaptive development in infants with autism spectrum disorder in the first 2 years of life. *Journal of Neurodevelopmental Disorders*, *7*, 24. <https://doi.org/10.1186/s11689-015-9117-6>
- Gable, P. A., & Harmon-Jones, E. (2008). Approach-motivated positive affect reduces breadth of attention. *Psychological Science*, *19*, 476–482. <https://doi.org/10.1111/j.1467-9280.2008.02112.x>
- Garon, N., Zwaigenbaum, L., Bryson, S., Smith, I. M., Brian, J., Roncadin, C., Vaillancourt, T., Armstrong, V., Sacrey, L.-A., & Roberts, W. (2016). Temperament and its association with autism symptoms in a high-risk population. *Journal of Abnormal Child Psychology*, *44*, 757–769. <https://doi.org/10.1007/s10802-015-0064-1>
- Gartstein, M. A., Knyazev, G. G., & Slobodskaya, H. R. (2005). Cross-cultural differences in the structure of infant temperament: United States of America (U.S.) and Russia. *Infant Behavior and Development*, *28*, 54–61. <https://doi.org/10.1016/j.infbeh.2004.09.003>
- Gartstein, M. A., & Rothbart, M. K. (2003). Studying infant temperament via the Revised Infant Behavior Questionnaire. *Infant Behavior and Development*, *26*, 64–86. [https://doi.org/10.1016/S0163-6383\(02\)00169-8](https://doi.org/10.1016/S0163-6383(02)00169-8)
- Goldsmith, H. H. (1996). Studying temperament via construction of the Toddler Behavior Assessment Questionnaire. *Child Development*, *67*, 218–235. <https://doi.org/10.1111/j.1467-8624.1996.tb01730.x>
- Gottesman, I. I., & Gould, T. D. (2003). The endophenotype concept in psychiatry: Etymology and strategic intentions. *American Journal of Psychiatry*, *160*, 636–645. <https://doi.org/10.1176/appi.ajp.160.4.636>
- Guthrie, W., Swineford, L. B., Nottke, C., & Wetherby, A. M. (2013). Early diagnosis of autism spectrum disorder: Stability and change in clinical diagnosis and symptom presentation. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *54*, 582–590. <https://doi.org/10.1111/jcpp.12008>
- Harmon-Jones, E., & Gable, P. A. (2009). Neural activity underlying the effect of approach-motivated positive affect on narrowed attention. *Psychological Science*, *20*, 406–409. <https://doi.org/10.1111/j.1467-9280.2009.02302.x>
- Joo, S. H., & Kim, E. S. (2019). Impact of error structure misspecification when testing measurement invariance and latent-factor mean difference using MIMIC and multiple-group confirmatory factor analysis. *Behavior Research Methods*, *51*, 2688–2699. <https://doi.org/10.3758/s13428-018-1124-6>
- Lemery-Chalfant, K., Doelger, L., & Goldsmith, H. H. (2008). Genetic relations between effortful and attentional control and symptoms of psychopathology in middle childhood. *Infant and Child Development: An International Journal of Research and Practice*, *17*, 365–385. <https://doi.org/10.1002/icd.581>
- Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford Press. <http://psycnet.apa.org/record/2013-17205-000>
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H. J., Leventhal, B. L., DiLavore, P. C., Pickles, A., & Rutter, M. (2000). The Autism Diagnostic Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, *30*, 205–223. <https://doi.org/10.1023/A:1005592401947>
- Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *24*, 659–685. <https://doi.org/10.1007/BF02172145>
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*, 519–530. <https://doi.org/10.1093/biomet/57.3.519>
- Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, *66*, 195–244. <https://doi.org/10.2466/pr0.1990.66.1.195>
- Miller, M., Iosif, A. M., Hill, M., Young, G. S., Schwichtenberg, A. J., & Ozonoff, S. (2017). Response to name in infants developing autism spectrum disorder: A prospective study. *Journal of Pediatrics*, *183*, 141–146. <https://doi.org/10.1016/j.jpeds.2016.12.071>
- Mullen, E. M. (1995). *Mullen Scales of Early Learning*. American Guidance Service Inc.
- Nyström, P., Gliša, T., Nilsson Jobs, E., Gredebäck, G., Charman, T., Johnson, M. H., Bölte, S., & Falck-Ytter, T. (2018). Enhanced pupillary light reflex in infancy is associated with autism diagnosis in toddlerhood. *Nature Communications*, *9*, 1–5. <https://doi.org/10.1038/s41467-018-03985-4>
- Ozonoff, S., Heung, K., Byrd, R., Hansen, R., & Hertz-Picciotto, I. (2008). The onset of autism: Patterns of symptom emergence in the first years of life. *Autism Research*, *1*, 320–328. <https://doi.org/10.1002/aur.53>
- Ozonoff, S., Iosif, A.-M., Baguio, F., Cook, I. C., Hill, M. M., Hutman, T., Rogers, S. J., Rozga, A., Sangha, S., Sigman, M., Steinfeld, M. B., & Young, G. S. (2010). A prospective study of the emergence of early behavioral signs of autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*, 256–266.e2. <https://doi.org/10.1097/00004583-201003000-00009>

- Ozonoff, S., Young, G. S., Belding, A., Hill, M., Hill, A., Hutman, T., Johnson, S., Miller, M., Rogers, S. J., Schwichtenberg, A. J., Steinfeld, M., & Iosif, A.-M. (2014). The broader autism phenotype in infancy: When does it emerge? *Journal of the American Academy of Child and Adolescent Psychiatry*, 53, 398–407. <https://doi.org/10.1016/j.jaac.2013.12.020>
- Ozonoff, S., Young, G. S., Carter, A., Messinger, D., Yirmiya, N., Zwaigenbaum, L., Bryson, S., Carver, L. J., Constantino, J. N., Dobkins, K., Hutman, T., Iverson, J. M., Landa, R., Rogers, S. J., Sigman, M., & Stone, W. L. (2011). Recurrence risk for autism spectrum disorders: A Baby Siblings Research Consortium study. *Pediatrics*, 128, e488–e495. <https://doi.org/10.1542/peds.2010-2825>
- Ozonoff, S., Young, G. S., Landa, R. J., Brian, J., Bryson, S., Charman, T., Chawarska, K., Macari, S. L., Messinger, D., Stone, W. L., Zwaigenbaum, L., & Iosif, A.-M. (2015). Diagnostic stability in young children at risk for autism spectrum disorder: A baby siblings research consortium study. *Journal of Child Psychology and Psychiatry*, 56, 988–998. <https://doi.org/10.1111/jcpp.12421>
- Paterson, S. J., Wolff, J. J., Elison, J. T., Winder-Patel, B., Zwaigenbaum, L., Estes, A., Pandey, J., Schultz, R. T., Botteron, K., Dager, S. R., Hazlett, H. C., & Piven, J. (2019). The importance of temperament for understanding early manifestations of autism spectrum disorder in high-risk infants. *Journal of Autism and Developmental Disorders*, 49(7), 2849–2863. <https://doi.org/10.1007/s10803-019-04003-2>
- Petersen, S., & Posner, M. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience*, 35, 73–89. <https://doi.org/10.1146/annurev-neuro-062111-150525>
- Pijl, M. K. J., Bussu, G., Charman, T., Johnson, M. H., Jones, E. J., Pasco, G., Buitelaar, J. K. (2019). Temperament as an early risk marker for autism spectrum disorders? A longitudinal study of high-risk and low-risk infants. *Journal of Autism and Developmental Disorders*, 49(5), 1825–1836. <https://doi.org/10.1007/s10803-018-3855-8>
- Posner, M. I., & Rothbart, M. K. (1991). Attentional mechanisms and conscious experience. In M. Rugg & A. D. Milner (Eds.), *The neuropsychology of consciousness* (pp. 91–111). Academic Press. <https://doi.org/10.1016/B978-0-12-498045-7.50010-4>
- Putnam, S. R. (2015). Temperament. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral science* (pp. 173–179). Elsevier.
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Rossee, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1–36. http://byrne.slab.net/classes/lavaan_materials/lavaanIntroduction4-9.pdf
- Rothbart, M. K. (1981). Measurement of temperament in infancy. *Child Development*, 52, 569–578. <https://doi.org/10.2307/1129176>
- Rothbart, M. K. (2012). Advances in temperament: History, concepts, and measures. In M. Zentner & R. L. Shiner (Eds.), *Handbook of temperament* (pp. 3–20). The Guilford Press.
- Rothbart, M. K., Ellis, L. K., Rueda, M., & Posner, M. I. (2003). Developing mechanisms of temperamental effortful control. *Journal of Personality*, 71, 1113–1143. <https://doi.org/10.1111/1467-6494.7106009>
- Rutter, M., Bailey, A., & Lord, C. (2003). *The Social Communication Questionnaire: Manual*. Western Psychological Services.
- Sacrey, L.-A., Bryson, S., Zwaigenbaum, L., Brian, J., Smith, I. M., Roberts, W., Szatmari, P., Vaillancourt, T., Roncadin, C., & Garon, N. (2018). The autism parent screen for infants: Predicting risk of autism spectrum disorder based on parent-reported behavior observed at 6–24 months of age. *Autism*, 22, 322–334. <https://doi.org/10.1177/1362361316675120>
- Shiner, R. L., Buss, K. A., McClowry, S. G., Putnam, S. P., Saudino, K. J., & Zentner, M. (2012). What is temperament now? Assessing progress in temperament research on the Twenty-Fifth Anniversary of Goldsmith et al (1987). *Child Development Perspectives*, 6, 436–444. <https://doi.org/10.1111/j.1750-8606.2012.00254.x>
- Shiner, R. L., & DeYoung, C. G. (2013). The structure of temperament and personality traits: A developmental perspective. In P. D. Zelazo (Ed.), *Oxford library of psychology. The Oxford handbook of developmental psychology, Vol. 2. Self and other* (pp. 113–141). Oxford University Press.
- Stallworthy, I. C., Lasch, C., Berry, D., Wolff, J. J., Pruett, J. R. Jr, Marrus, N., Swanson, M. R., Botteron, K. N., Dager, S. R., Estes, A. M., Hazlett, H. C., Schultz, R. T., Zwaigenbaum, L., Piven, J., & Elison, J. T. for the IBIS Network. (2022). Variability in responding to joint attention cues in the first year is associated with autism outcome. *Journal of the American Academy of Child and Adolescent Psychiatry*, 61, 413–422. <https://doi.org/10.1016/j.jaac.2021.03.023>
- Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments; applications in the substance use domain. In K. J. Bryant, S. G. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). American Psychological Association. <https://doi.org/10.1037/10222-009>
- Wolff, J. J., Botteron, K. N., Dager, S. R., Elison, J. T., Estes, A. M., Gu, H., Hazlett, H. C., Pandey, J., Paterson, S. J., Schultz, R. T., Zwaigenbaum, L., & Piven, J. (2014). Longitudinal patterns of repetitive behavior in toddlers with autism. *Journal of Child Psychology and Psychiatry*, 55, 945–953. <https://doi.org/10.1111/jcpp.12207>
- Yu, C., & Smith, L. B. (2016). The social origins of sustained attention in one-year-old human infants. *Current Biology*, 26, 1235–1240. <https://doi.org/10.1016/j.cub.2016.03.026>
- Zwaigenbaum, L., Bryson, S. E., Brian, J., Smith, I. M., Roberts, W., Szatmari, P., Roncadin, C., Garon, N., & Vaillancourt, T. (2016). Stability of diagnostic assessment for autism spectrum disorder between 18 and 36 months in a high-risk cohort. *Autism Research*, 9, 790–800. <https://doi.org/10.1002/aur.1585>
- Zwaigenbaum, L., Bryson, S. E., Brian, J., Smith, I. M., Sacrey, L., Armstrong, V., Roberts, W., Szatmari, P., Garon, N., & Vaillancourt, T., & Roncadin, C. (2021). Assessment of autism symptoms from 6 to 18 months of age using the autism observational scale for infants in a prospective high-risk cohort. *Child Development*, 92, 1187–1198. <https://doi.org/10.1111/cdev.13485>
- Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., & Szatmari, P. (2005). Behavioral manifestations of autism in the first year of life. *International Journal of Developmental Neuroscience*, 23, 143–152. <https://doi.org/10.1016/j.ijdevneu.2004.05.001>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Sung, S., Fenoglio, A., Wolff, J. J., Schultz, R. T., Botteron, K. N., Dager, S. R., Estes, A. M., Hazlett, H. C., Zwaigenbaum, L., Piven, J., & Elison, J. T. (2022). Examining the factor structure and discriminative utility of the Infant Behavior Questionnaire–Revised in infant siblings of autistic children. *Child Development*, 93, 1398–1413. <https://doi.org/10.1111/cdev.13781>