7-1-2008

Homotypic and heterotypic continuity of fine-grained temperament during infancy, toddlerhood, and early childhood

Samuel P. Putnam
Bowdoin College

Mary K. Rothbart
University of Oregon

Maria A. Gartstein
Washington State University

Infant and Child Development
Inf. Child. Dev. **17**: 387–405 (2008)
Published online in Wiley InterScience
(www.interscience.wiley.com). DOI: 10.1002/ICD.582

Homotypic and Heterotypic Continuity of Fine-grained Temperament during Infancy, Toddlerhood, and Early Childhood

Samuel P. Putnam[a,*], Mary K. Rothbart[b] and Maria A. Gartstein[c]

[a]*Department of Psychology, Bowdoin College, College Station, Brunswick, ME, USA*
[b]*Department of Psychology, University of Oregon, OR, USA*
[c]*Department of Psychology, Washington State University, USA*

Longitudinal continuity was investigated for fine-grained and factor-level aspects of temperament measured with the Infant Behaviour Questionnaire-Revised (IBQ-R), Early Childhood Behaviour Questionnaire (ECBQ), and Children's Behaviour Questionnaire (CBQ). Considerable homotypic continuity was found. Convergent and discriminant validity of the measures was supported, as all fine-grained dimensions exhibited stability across adjacent measurement periods, and all scales found on both the ECBQ and CBQ were most highly correlated with their equivalent scales. At the factor level, Surgency and Negative Affect factors were stable across all time points, and Effortful Control/Regulatory Capacity was stable across adjacent time periods. High-Intensity Pleasure, Activity Level, and Impulsivity contributed strongly to continuity of Surgency, and Sadness, Frustration, and Falling Reactivity played strong roles in the continuity of Negative Affect. Heterotypic continuity was also found. High levels of Infant Surgency predicted high toddler Effortful Control, whereas high toddler Surgency predicted low Effortful Control in preschoolers. Infant Surgency dimensions especially predicted Toddler Attention Shifting and Low-Intensity Pleasure, and toddler Activity Level was most closely associated with later deficits in Effortful Control. Inverse relations were also obtained between Negative Affect and Effortful Control, with substantial negative connections between toddler Negative Affect and preschool Attention

*Correspondence to: Sam Putnam, Department of Psychology, Bowdoin College, 6900 College Station, Brunswick, ME 04011, USA. E-mail: sputnam@Bowdoin.edu

Key words: temperament; personality; longitudinal; continuity; infants; toddlers

The majority of longitudinal research on temperament has focused on single aspects of behaviour (e.g. Kagan, 1994), traits with clear genetic underpinnings (e.g. Buss and Plomin, 1975), infant 'difficultness' (e.g. Bates, Freeland, & Lounsbury, 1979), or the nine characteristics identified in the landmark work of Thomas and Chess (McDevitt & Carey, 1978; Thomas, Chess, & Birch, 1968). In contrast, recent questionnaire research has established a more comprehensive array of traits. The current study investigates the longitudinal stability of attributes assessed by fine-grained temperament questionnaires designed for use with infants (Infant Behaviour Questionnaire—Revised [IBQ-R]; Gartstein & Rothbart, 2003), toddlers (Early Childhood Behaviour Questionnaire [ECBQ]; Putnam, Gartstein, & Rothbart, 2006), and young children (Children's Behaviour Questionnaire [CBQ]; Rothbart, Ahadi, Hershey, & Fisher, 2001).

A combination of empirical findings and theoretical perspectives informed the creation of these instruments. Item-level factor analyses (Sanson, Prior, Garino, & Oberklaid, 1987) of instruments designed to measure the nine dimensions derived by Thomas and Chess revealed a lack of coherence among items measuring some dimensions, and revealed alternative dimensions. Growing recognition of distinctions between primary emotions (e.g. Izard, 1977) prompted the creation of scales measuring discrete aspects of emotionality, in contrast to broader measures of 'mood' (e.g. McDevitt & Carey, 1978) included in earlier measures. Optimal level of arousal theories (e.g. Bell, 1974) emphasizing the importance of situational context led to consideration of dimensions that addressed child behaviour in high- and low-intensity situations. Finally, research on the development of attentional self-regulatory systems, reflected in Rothbart's (Rothbart & Bates, 2006; Rothbart & Derryberry, 1981) definition of temperament as involving individual differences in reactivity and regulation, suggested the need for scales assessing children's capabilities for attentional, emotional, and behavioural control.

A fine-grained perspective of temperament confers at least two advantages over other approaches. First, more precise investigations of stability, instability, and relations with other variables are possible in fine-grained analyses of child temperament. Second, scale-level factor analysis examining interrelations of traits can be used to derive information regarding the hierarchical structure of temperament, an issue that has been largely neglected in the literature regarding children (Caspi, 1998). Analyses of fine-grained instruments developed for use from infancy through adulthood have revealed factors of Surgency, Negative Affectivity, and Orienting/Regulatory Capacity or Effortful Control that bear similarity to constructs that have emerged in models of adult personality (Putnam, Ellis, & Rothbart, 2001; see Table 1 for definitions and factor affiliation of IBQ-R, ECBQ, and CBQ scales). The current report examines the stability of these higher order factors, utilizing discrete, narrowly defined traits to determine more precise contributions to continuity among these factors.

Developmental psychologists refer to at least five types of continuity in longitudinal research (Caspi, 1998), three of which are relevant to the current study. The *structural continuity* of temperament concerns relations between sets of variables and is indicated by the largely similar structures that have emerged from factor analyses of the IBQ-R, ECBQ, and CBQ. More pertinent to the pur-

DOI: 10.1002/icd

Table 1. Labels, definitions, and factor affiliation of IBQ-R, ECBQ, and CBQ scales

Factor	Label	Definition	IBQ-R	ECBQ	CBQ
Surgency					
	Activity Level	Gross motor activity, including rate and extent of movement	Sur	Sur	Sur
	High-Intensity Pleasure	Pleasure or enjoyment related to high stimulus intensity, rate, complexity, novelty, and incongruity	Sur	Sur	Sur
	Positive Anticipation	Excitement in the anticipation of pleasurable activities. Includes rapid approach in IBQ-R	Sur	Sur	Sur
	Impulsivity	Speed of response initiation		Sur	Sur
	Sociability	Seeking and taking pleasure in interactions with others		Sur	
	Vocal Reactivity	Vocalization during daily activities	Sur		
Negative Affect					
	Falling Reactivity	Rate of recovery from peak distress, excitement, or general arousal; ease of falling asleep. Also includes soothability items in ECBQ and CBQ	Neg	Neg	Neg
	Fear	Negative affect related to anticipated pain, distress and/or threat. Includes startle and reactions to novelty social stimuli in IBQ-R	Neg	Neg	Neg
	Frustration/ distress to limitations	Negative affect related to confinement, interruption of ongoing tasks or goal blocking	Neg	Neg	Neg
	Sadness	Negative affect, tearfulness or lowered mood related to physical state, disappointment, loss, and/or response to other's suffering	Neg	Neg	Neg
	Discomfort	Negative affect related to sensory qualities of stimulation, including intensity, rate or complexity of light, sound, and texture		Neg	Neg
	Motor activation	Repetitive small motor movements; fidgeting		Neg	
Orienting/Regulatory Capacity/Effortful Control					
	Attention/Duration of Orienting	Capacity to sustain attention on an object or task. Also includes ability to shift attention on CBQ	Eff	Eff	Eff
	Low-intensity Pleasure	Pleasure or enjoyment related to low stimulus intensity, rate, complexity, novelty, and incongruity	Eff	Eff	Eff
	Affiliation/ Cuddliness	Desire for, and pleasure in, warmth and closeness with others, independent of shyness or extraversion	Eff	Eff	
	Inhibitory Control	Capacity to suppress inappropriate actions or responses. Includes planning capabilities in CBQ		Eff	Eff
	Attentional Shifting	Ability to transfer attentional focus from one object or task to another		Eff	
	Soothability	Reduction of fussing, crying, or distress when soothing techniques are used by the caretaker	Eff		
Inconsistent factor affiliation					
	Perceptual Sensitivity	Detection of slight, low-intensity stimuli from the external environment	Sur	Neg	Eff
	Shyness	Slow or inhibited approach and/or discomfort in social situations		Neg	Sur
	Smiling and Laughter	Positive affect in response to changes in stimulus intensity, rate, complexity, and incongruity	Sur		Eff

DOI: 10.1002/icd

pose of this study are constructs of *differential continuity*, referring to the maintenance of an individual's relative position in a group, and the concept of *coherence*. According to Caspi (1998, pp. 349–350), consideration of coherence 'enlarges the definition of continuity to include heterotypic continuity—continuity of an inferred genotypic attribute presumed to underlie diverse phenotypic behaviors'.

The majority of studies concerning differential continuity of temperament and personality have concerned homotypic continuity—continuity of similar behaviours over time. Several scales contained on all three measures used in the current study represent this aspect of continuity. Although individual items of the IBQ-R, ECBQ, and CBQ differ in their eliciting contexts, the behavioural descriptors assessed for corresponding scales are largely consistent. For instance, IBQ-R Activity Level items describe behaviour while being dressed or fed, whereas CBQ Activity Level items refer to situations such as playing outside or moving from room to room, but child behaviours assessed through items on both instruments reflect large motor movements.

Whereas homotypic continuity is most likely to be evident after puberty, Kagan (1969) has argued that, due to rapid developmental changes in the early years of life, a great deal of continuity during childhood will be heterotypic. Examples of heterotypic continuity are evident in the literature. For instance, Kagan, Snidman, and Arcus (1998) found 4-month-old infants who reacted to novel stimuli with high amounts of Negative Affect and activity were likely to avoid interaction with peers at 4 years of age. Heterogeneity is similarly evident in items used to assess temperament dimensions through parent report. Whereas fear is measured in the IBQ-R with items asking how often the baby startled at a sudden change in body position or how often crying was observed in the presence of unfamiliar adults, the same dimension is tapped in the CBQ with items that do not presume crying or startle reactions (e.g. 'My child is afraid of fire').

Heterogeneity should be at a maximum for traits that show substantial development over the time period in question. Among the attributes measured with our fine-grained temperament measures, the most profound changes are apparent for dimensions assessing attentional and self-regulatory aspects of child functioning that show dramatic transformations between birth and early childhood (Posner & Rothbart, 2007). To a large extent, attention in infancy is reactive in nature, becoming more effortfully controlled with maturation (Posner & Rothbart, 2007; Ruff & Rothbart, 1996). As such, the IBQ-R includes a scale measuring Duration of Orienting that refers to tendencies to focus on a single object for long periods, whereas Attentional Focusing scales on the ECBQ and CBQ contain items referring to the ability to willfully maintain attention in the face of distraction. Because behavioural control is largely achieved with adult support during infancy but becomes more autonomous over time (Kopp, 1982), the ECBQ and CBQ, but not the IBQ-R, include scales measuring the child's Inhibitory Control.

The heterogeneity of behaviours reflecting common underlying processes suggests the value of aggregating across narrowly delineated temperament characteristics. Common neural processes presumably unite the differentiated traits constituting temperament factors, although different circuits may contribute to different negative emotions (Gray, 1991). For example, recent theories converge on the anterior cingulate gyrus and lateral prefrontal cortex as playing an organizing role in attentional and behavioural control, and on dopamine activity in relation to Extraversion/Surgency (Posner & Rothbart, 2007; Rothbart & Bates, 2006). Although the current research makes use of fine-grained attributes, with the prediction that scales loading on a given factor will predict later traits associated with that factor

DOI: 10.1002/icd

(e.g. Activity Level will predict later Activity Level, Positive Anticipation, High-Intensity Pleasure, Impulsivity, and Sociability), we expected more substantial stability at the factor level, since aggregation may allow more robust measurement of individual differences rooted in biological systems.

Developmental processes underlying continuity, however, may be evident in ways beyond factor-level stability. Indeed, common underlying mechanisms may contribute to multiple behavioural traits, and dynamic relations can implicate one system in the development of another. A well-documented example concerns anger. As demonstrated by the consistent loading of Frustration/Anger on Negative Affect factors, a common process links anger to other negative affects. Anger, however, is also tied to aspects of Surgency. Rothbart, Derryberry, and Hershey (2000) found prediction from infant Activity Level to 7-year frustration, and Putnam and Stifter (2005) reported high levels of externalizing behaviour in highly surgent 'exuberant' toddlers. Exemplifying the interplay between different developmental processes, Stifter and Spinrad (2002) have reported data suggesting that excess crying during early infancy may hamper the development of self-regulatory skills.

Recent findings from Komsi *et al*. (2006) are particularly informative in regards to expectations for longitudinal relations across factors in the current study. Infant Negative Affectivity, measured using the IBQ (Rothbart, 1981; the predecessor to the IBQ-R) was predictive of CBQ Negative Affectivity at age 5. An infant 'Positive Affectivity' factor consisting of Activity Level, Smiling and Laughter, Soothability, and Duration of Orienting was predictive not only of later Surgency/Extraversion, but also of Effortful Control. Scale-level analyses indicated that prediction to later Surgency was particularly strong for infant Activity Level, whereas Smiling and Laughter and Duration of Orienting were more closely related to preschool Effortful Control, suggesting that early forms of Positive Affectivity may either be markers of self-regulatory capacity or may allow for the development of self-regulation (Komsi *et al*., 2006).

Komsi *et al*. (2006) suggested a limitation of their study was the lack of an intermediate assessment of temperament between infancy and the preschool period, a shortcoming addressed in the current study. Although we expected to conceptually replicate the prior findings through positive correlations between infant Surgency to later Effortful Control, there was reason to expect that this relation may not hold, or might even be negative in direction, from toddlerhood to early childhood. Conceptually, the metaphor of the approach-oriented systems underlying Surgency functioning as an 'accelerator' that interacts with the 'brakes' of Effortful Control (Derryberry & Rothbart, 1997) suggests the two may be inversely related over time. Empirically, variable patterns of concurrent relations between these traits have been observed. Whereas Gartstein and Rothbart (2003) obtained a significant positive correlation between Surgency and ORC during infancy, Putnam *et al*. (2006) found Surgency and Effortful Control to be uncorrelated among toddlers, and Putnam, Rothbart, and Gartstein (2004) reported a significant negative correlation between the two factors from the CBQ. The direction of relations between other factors has been more consistent across the instruments. For all three measures, negative correlations have been consistently obtained between Negative Affectivity and Effortful Control/Regulatory Capacity, and on both the IBQ-R and CBQ, positive correlations existed between Surgency and Negative Affectivity. We expected these findings to be conceptually replicated in our longitudinal analyses.

We agree with Caspi (1998) and Fox and Henderson (1999) that examinations of heterotypic continuity will be most productive when directed by theory, and

DOI: 10.1002/icd

believe that attempts at replication are similarly useful. However, exploratory analyses can also be used effectively to aid in generating theory and/or revising existing principles. Due to the relatively recent development of the fine-grained approach, little theory or data concerning longitudinal connections between discrete traits have been generated to date, and thus, an inductive approach seems warranted at this time. In addition to testing the hypothesized factor-level relations described above, we also report exploratory longitudinal correlational analyses to explain patterns of continuity demonstrated between factors.

METHODS

Participants

A community sample of primary caregivers (90% mothers) of 361 (181 female) infants from a small city in the northwest US was initially recruited and administered the IBQ-R. Infants were equally distributed across three age groups: 3–6, 6–9, and 9–12 months. Participants were recruited by telephone on the basis of birth announcements published in the local paper. Nearly all subjects recruited were Caucasian and married. The sample was primarily middle class, with an average family income of $41,798 (S.D. = 19, 154) and average Revised Duncan Sociometric Index (Stevens & Featherman, 1981) of 54.12 (S.D. = 19.20). Primary caregivers' average age was 31.10 (S.D. = 5.30) years, and they had completed an average of 14.51 (S.D. = 2.40) years of education. Additional details regarding subject recruitment and characteristics can be found in Gartstein and Rothbart (2003).

From this initial group of 361, primary caregivers of 248 (125 female) children completed the ECBQ. Comparisons of those who did and did not participate at this time revealed no significant differences in education level, yearly income, respondent age, or any of the temperament variables assessed during infancy. To supplement this sample, families of 68 (39 female) additional children were recruited. Demographic information was not gathered from these 68 families, but the similarity in recruitment strategy allows some degree of confidence that they represent the same population as the original sample. The total sample of 316 children (164 female) was roughly equally distributed across three age groups: 18–22 ($n = 103$), 23–26 ($n = 110$), and 27–32 months ($n = 104$). See Putnam *et al.* (2006) for further details.

Of the 316 participants who completed the ECBQ, 187 (101 female) completed the CBQ. The Time 3 subsample was normally distributed with respect to child age ($M = 49.51$, S.D. = 4.81), and included 142 children (75 female) whose primary caregiver had completed both the IBQ-R and ECBQ, and 45 (26 female) whose parent only completed the ECBQ. When demographics, infant temperament, and toddler temperament of participants who took part in all three assessments were compared with those who dropped out of the study between infancy and childhood, only 2 of 36 *t*-tests were significant. Infants who remained in the study were rated by parents as being higher in IBQ-R Sadness, $t(357) = 2.54$, $p < 0.05$ and lower in IBQ-R Soothability, $t(358) = 2.04$, $p < 0.05$, than infants who were lost to attrition.

Procedure

Parents of infants between 3 and 12 months of age whose births were reported in the local newspaper were contacted by phone. Potential subjects were told that their participation would involve completing a consent form, the IBQ-R (interim

DOI: 10.1002/icd

version) and a demographic form, that completing the forms would take approximately 1 h, and that they would receive a check for $5 for their participation, as well as a self-addressed, stamped envelope to return the completed forms. Of 476 parents contacted, 361 (76%) completed the measures. See Gartstein and Rothbart (2003) for additional details. When children were between the ages of 18 and 32 months, parents were again contacted by phone and mail, then mailed a consent form, ECBQ (interim version), a self-addressed, stamped envelope, and either a check for $5 or a $5 gift certificate to a local toy store. See Putnam *et al*. (2006) for details. When the children whose parents had completed the ECBQ were between 37 and 59 months old, parents were sent the CBQ, a self-addressed, stamped envelope and a $5 check. Three weeks after this mailing, parents who had not returned completed forms were contacted by phone, and three weeks after these calls, a final reminder letter was sent. Of the 130 families who dropped out between the ECBQ and CBQ assessments, 11 had relocated and no forwarding information was available, 37 declined to participate when contacted by phone, 13 failed to return forms despite agreeing to participate when contacted by phone, and 69 could not be reached by phone.

Materials

The IBQ-R, ECBQ, and CBQ are parent-report instruments containing items referring to child behaviour in commonly occurring situations. For each item, parents are asked to rate the child on a 7-point Likert-type scale. The IBQ-R and ECBQ items are phrased in the form of questions about the child's behaviour in a given context during the past 1 or 2 weeks (e.g. 'When being carried in the past week, how often did the baby push against you until put down?'), and the ratings refer to frequency of behaviour (never, very rarely, less than half the time, half the time, more than half the time, almost always, always). The CBQ items are statements describing child behaviour (e.g. 'My child gets angry when told he or she needs to go to bed'), and the ratings refer to the degree to which the statement accurately describes the child (extremely untrue, quite untrue, slightly untrue, neither true nor untrue, slightly true, quite true, extremely true).

The sample used in the current study contributed to the refinement of the IBQ-R and ECBQ, such that participants received interim versions of the instruments containing some items that were ultimately discarded due to statistical and conceptual considerations (see Gartstein & Rothbart, 2003; Putnam *et al*., 2006). The interim IBQ-R contained 251 items, 67 of which were eliminated prior to creating scale scores. The interim ECBQ contained 267 items, 66 of which were removed.

Table 1 contains the labels and definitions for the scales included on the three instruments. Several scales, such as High-Intensity Pleasure and Frustration/Distress to Limitations, are used in each of the measures, while others appear on only one or two of the instruments. Cronbach's alphas for the 14 scales of the IBQ-R ranged from 0.70 to 0.90. Alphas for the ECBQ ranged from 0.61 to 0.89. CBQ scale alphas ranged from 0.61 to 0.94. As discussed in the introduction, factor analyses of each instrument had identified three-factor solutions. Scores corresponding to these factors were created by performing principal axis factor analyses on scale scores, rotating extracted factors using the Oblimin algorithm, and saving the variables created by the regression method (Norusis, 1994).

DOI: 10.1002/icd

RESULTS

Previously reported findings (Gartstein & Rothbart, 2003; Putnam *et al*., 2006; Rothbart *et al*., 2001) using the samples and/or measures used in the current study mirror findings from other studies (see Else-Quest, Hyde, Goldsmith, & Van Hulle, 2006) in demonstrating higher scores on dimensions of Surgency in males and Effortful Control in females. Because there is no apparent rationale for expecting gender differences in patterns of stability (Komsi *et al*., 2006), we did not conduct separate analyses for males and females. In addition, parent ratings of several dimensions assessed with these questionnaires increase with child age (Gartstein & Rothbart, 2003; Goldsmith, 1996; Putnam *et al*., 2006). To control for sex and age influences, these variables were residualized from all temperament scores prior to substantive analyses. The analyses below were conducted using these residual scores. We first present continuity findings at the factor level, before defining these results more thoroughly by considering fine-grained stability.

Continuity of Temperament

Factor scores

As shown in Table 2, with one exception, all cross-age correlations for Surgency and Negative Affect were significant and moderate to large in magnitude. ECBQ Effortful Control (EFF) was related to both the IBQ-R Orienting/Regulatory Capacity (ORC) factor, and CBQ EFF, but the infant and childhood factors were not significantly related. Between infancy and toddlerhood, SUR and ORC-EFF were positively correlated, with infant SUR additionally predicting preschool EFF. This pattern was reversed between toddlerhood and preschool age, with ECBQ SUR negatively correlated with CBQ EFF. In addition, negative correlations were obtained between IBQ-R ORC and CBQ NEG and between ECBQ NEG and CBQ EFF. No correlations between SUR and NEG were significant.

Simultaneous multiple regression was first employed to examine whether SUR and ORC in infancy each predicted unique variance in ECBQ SUR and EFF. The model for ECBQ SUR was significant, $F(2, 243) = 18.49$, $p < 0.01$, $R^2 = 0.13$, with IBQ-R SUR, $\beta = 0.38$, $p < 0.01$, but not IBQ-R ORC, $\beta = -0.03$, n.s., accounting for unique variance. The model for ECBQ EFF was also significant, $F(2, 243) = 28.27$, $p < 0.01$, $R^2 = 0.19$, with both IBQ-R SUR, $\beta = 0.31$, $p < 0.01$, and IBQ-R ORC, $\beta = 0.17$, $p < 0.01$, making unique predictions.

Table 2. Cross-age correlations between factors

		ECBQ factors			CBQ factors		
		SUR	NEG	EFF	SUR	NEG	EFF
IBQ-R factors	SUR	0.36**	0.08	0.41**	0.25**	0.04	0.16*
	NEG	0.05	0.36**	−0.10	0.06	0.36**	0.05
	ORC	0.17**	−0.07	0.34**	0.10	−0.20*	0.05
ECBQ factors	SUR				0.59**	−0.01	−0.16*
	NEG				0.10	0.49**	−0.17*
	EFF				−0.13#	−0.11	0.58**

Note: $N = 246$ for IBQ-R to ECBQ, 186 for ECBQ to CBQ, and 141 for IBQ-R to CBQ. SUR, Surgency; NEG, Negative Affectivity; ORC, Orienting/Regulatory Capacity; EFF, Effortful Control. # $p < 0.10$, * $p < 0.05$, ** $p < 0.01$.

DOI: 10.1002/icd

Simultaneous multiple regression was also used to examine unique variance in CBQ EFF predicted by the three ECBQ factors. The model was significant, $F(3, 182) = 32.91$, $p < 0.01$, $R^2 = 0.35$. ECBQ EFF accounted for unique variance, $\beta = 0.57$, $p < 0.01$, the coefficient for ECBQ SUR was marginally significant, $\beta = -0.11$, $p < 0.10$, but ECBQ NEG was not significant, $\beta = -0.01$, n.s.

Hierarchical multiple regression was utilized to assess the mediating role of toddler temperament in explaining infant–preschool relations. As expected, when entered alone, IBQ-R SUR predicted CBQ SUR, $\beta = 0.22$, $p < 0.01$, model $F(1, 138) = 6.82$, $p < 0.01$, $R^2 = 0.05$. Suggesting mediation, when ECBQ SUR was also included in a subsequent model, $F(2, 137) = 27.86$, $p < 0.01$, $R^2 = 0.29$, the coefficient for IBQ-R SUR was no longer significant, $\beta = -0.03$, n.s., and the coefficient for ECBQ SUR was significant, $\beta = 0.55$, $p < 0.01$. In contrast, IBQ-R NEG was not completely mediated by ECBQ in the prediction of CBQ NEG. When entered alone, the coefficient for IBQ-R was significant, $\beta = 0.36$, $p < 0.01$, model $F(1, 138) = 20.35$, $p < 0.01$, $R^2 = 0.36$. When ECBQ NEG was added to the equation, the IBQ-R coefficient decreased, but remained significant, $\beta = 0.20$, $p < 0.05$, model $F(2, 137) = 21.22$, $p < 0.01$, $R^2 = 0.49$, ECBQ NEG $\beta = 0.36$, $p < 0.01$. Finally, the relation between IBQ-R SUR and CBQ EFF was mediated through ECBQ EFF (because the direction of relations between IBQ-R SUR and ECBQ SUR was opposite of those between ECBQ SUR and CBQ EFF, it was not considered as a mediator): IBQ-R accounted for unique variance in CBQ EFF when entered alone, $\beta = 0.20$, $p < 0.05$, model $F(1, 138) = 5.79$, $p < 0.05$, $R^2 = 0.20$, but not when ECBQ EFF was also included, $\beta = 0.01$, n.s., model $F(2, 137) = 32.89$, $p < 0.01$, $R^2 = 0.32$, ECBQ EFF $\beta = 0.57$, $p < 0.01$.

Fine-grained traits

A disadvantage to the fine-grained approach to temperament concerns the elevated potential for Type 1 error. When multiple statistical tests are performed, a high likelihood exists for falsely rejecting the null hypotheses and concluding (inaccurately) that a relationship exists between two variables. This problem is typically handled by employing the Bonferroni or similar procedures, but with very large numbers of tests, the Bonferroni correction is considered by many to be excessively conservative (Sankoh, Huque, & Dubey, 1997). We have employed an alternative procedure recommended by Uitenbroek (1997), dividing our sample into two random halves and only reporting heterotypic correlations that exceeded the traditionally accepted $p < 0.05$ in both samples, such that significance would be obtained at random only once in 400 tests. The correlations reported in Tables 3 and 4 were significant in both random subsamples, but to most accurately represent the relations, we have presented coefficients that were calculated using the entire sample. In the text below regarding heterotypic continuity, we do not describe all significant correlations, instead focusing on those we feel best explain our factor-level results.

Homotypic continuity

As shown in bold type in Table 3, convergent validity between the IBQ-R to the ECBQ was demonstrated through significant correlations for each of the 11 constructs included on both measures. Discriminant validity was demonstrated for 5 of the 11 scales (Activity Level, Positive Anticipation, Falling Reactivity, Perceptual Sensitivity, and Cuddliness) that were most highly correlated with their equivalent scale.

DOI: 10.1002/icd

Table 3. Correlations between IBQ-R and ECBQ scales

IBQ-R dimensions		ECBQ dimensions																	
		Surgency					Negative Affectivity								Effortful Control				
		Hip	Act	Imp	Soc	Pan	Fru	Sad	Dis	Mot	Fea	Fal	Shy	Psn	Atf	Inh	Ats	Lip	Cud
Surgency	Hip	**0.30**												0.25			0.26	0.36	
	Act		**0.32**							0.19									
	Pan					**0.32**								0.29			0.30	0.27	
	Smi	0.19										0.20		0.31	0.26		0.30	0.38	
	Voc		0.22		0.26									0.35				0.34	
	Psen													**0.45**	0.24		0.32	0.20	
Negative Affectivity	Fru						**0.22**					−0.25							
Orienting/Regulatory Capacity	Sad						0.25	**0.24**	0.18										
	Fall											**0.30**							
	Fear										**0.23**		0.31						
	Atfo													0.28	**0.23**		0.23	0.31	
	Lip																	**0.34**	
	Soo																	0.24	
	Cud																	0.25	**0.36**

Note: $N = 248$. For all correlations presented: $p < 0.05$ in two random subsamples. Bold values represent homotypic continuity coefficients. Hip = High-Intensity Pleasure; Act = Activity Level; Imp = Impulsivity; Soc = Sociability; Pan = Positive Anticipation; Fru = Frustration/Distress to Limitations; Sad = Sadness; Dis = Discomfort; Mot = Motor Activation; Fea = Fear; Fal = Falling Reactivity; Shy = Shyness; Psn = Perceptual Sensitivity; Atf = Attention Focusing/Duration of Orienting; Inh = Inhibitory Control; Ats = Attention Shifting; Lip = Low-Intensity Pleasure; Cud = Cuddliness; Voc = Vocal Reactivity; Smi = Smiling and Laughter; Soo = Soothability.

Inf. Child. Dev. **17**: 387–405 (2008)
DOI: 10.1002/icd

As shown in bold type in Table 4, all 14 scales included on both the ECBQ and CBQ demonstrated longitudinal stability from toddlerhood to early childhood. Furthermore, all scales were most highly correlated with their equivalent scale.

Only 6 of the 11 discrete traits measured with both the IBQ-R and CBQ demonstrated significant stability, and for only 2 of these, Perceptual Sensitivity and Distress to Limitations/Frustration were the highest correlations obtained with equivalent scales. Stability coefficients for Activity Level, Approach/Positive Anticipation, Smiling and Laughter, Distress to Limitations/Frustration, Sadness, and Perceptual Sensitivity were significant, rs(140) = 0.22, 0.23, 0.26, 0.29, 0.18, 0.23, ps < 0.05, whereas High-Intensity Pleasure, Fear, Falling Reactivity, Duration of Orienting/Attentional Focusing, and Low-Intensity Pleasure were not, rs(140) = 0.15, 0.13, 0.13, 0.08, 0.08, ps > 0.05.

Within-factor heterotypic continuity

As shown in Table 3, infant to toddler continuity of Surgency was best explained by homotypic relations. From toddlerhood to childhood, Table 4 reveals moderately large correlations among ECBQ and CBQ High-Intensity Pleasure, Activity Level, and Impulsivity. Further suggesting the centrality of Activity Level and Impulsivity in Surgency, IBQ-R Activity Level also predicted CBQ Impulsivity in both random subsamples, $r(142) = 0.25$, $p < 0.01$, in the full sample. In addition, IBQ-R Vocal Reactivity was predictive of CBQ Smiling and Laughter, full sample $r(142) = 0.35$, $p < 0.01$.

Homotypic relations also accounted for the majority of significant links between infant and toddler Negative Affectivity. From toddlerhood to the preschool period, however, substantial overlapping longitudinal prediction is evident among Frustration, Sadness, and Falling Reactivity. Only one heterotypic Negative Affectivity IBQ-R to CBQ correlation was significant in both random subsamples: IBQ-R Sadness predicted CBQ Discomfort, full sample $r(142) = 0.28$, $p < 0.01$.

Intriguingly, Low-Intensity Pleasure made particularly salient contributions to the stability of the regulatory factor, as all scales from IBQ-R ORC predicted ECBQ Low-Intensity Pleasure, and all scales from ECBQ Effortful Control predicted CBQ Low-Intensity Pleasure. From the toddler to preschool periods, ECBQ Attention Focusing and Inhibitory Control correlated with the same dimensions of the CBQ. No scales from the regulatory factors of the IBQ-R and CBQ were significantly correlated in both subsamples.

Cross-factor heterotypic continuity

As shown in Table 3, the fascinating positive correlation between infant Surgency and toddler Effortful Control is largely explained by connections between IBQ-R Surgency dimensions and ECBQ Attention Shifting and Low-Intensity Pleasure. ECBQ Perceptual Sensitivity, which loads substantially on Effortful Control, was also predicted by most IBQ-R Surgency scales. In contrast, the negative relation between these factors from the ECBQ to CBQ is largely due to negative relations between toddler Activity Level and preschool Attention Focusing and Inhibitory Control. It is of note that deficits in toddler Inhibitory Control are similarly related to excess Activity Level and Impulsivity in preschool. No scale-level correlations between IBQ-R Surgency and CBQ Effortful Control were significant in both subsamples.

DOI: 10.1002/icd

Table 4. Correlations between ECBQ and CBQ scales

ECBQ dimensions		CBQ dimensions														
		Surgency						Negative Affectivity					Effortful Control			
		Hip	Act	Imp	Shy	Smi	Pan	Fru	Sad	Dis	Fea	Fal	Atf	Inh	Lip	Psn
Surgency	Hip	**0.44**	0.34	0.31						−0.27						
	Act	0.36	**0.56**	0.40			0.32	0.26					−0.33	−0.36		
	Imp	0.29	0.32	**0.36**			0.29									
	Soc			0.31	−0.28											
	Pan		0.23				**0.40**									
Negative Affectivity	Fru		0.29				0.32	**0.52**	0.38			−0.36	−0.34	−0.32		
	Sad							0.30	**0.38**			−0.21	−0.26			
	Dis									**0.29**						
	Mot											−0.26		−0.28	−0.23	
	Fea										**0.41**					
	Fal							−0.31	−0.24			**0.44**	0.30	0.32		
	Shy				**0.49**						0.24					
	Psn															**0.41**
Effortful Control	Atf												**0.40**	0.26	0.28	
	Inh		−0.32	−0.32				−0.38				0.32	0.39	**0.51**	0.30	
	Ats											0.25			0.31	
	Lip					0.32									**0.40**	
	Cud														0.38	

Note: $N = 187$. For all correlations presented: $p < 0.05$ in two random subsamples. Bold values represent homotypic continuity coefficients. Hip = High-Intensity Pleasure; Act = Activity Level; Imp = Impulsivity; Shy = Shyness; Smi = Smiling and Laughter; Pat = Positive Anticipation; Fru = Frustration; Sad = Sadness; Dis = Discomfort; Fea = Fear; Fal = Falling Reactivity; Atf = Attention Focusing; Inh = Inhibitory Control; Lip = Low-Intensity Pleasure; Psn = Perceptual Sensitivity; Soc = Sociability; Mot = Motor Activation; Ats = Attention Shifting; Cud = Cuddliness.

DOI: 10.1002/icd

Finally, inspection of Table 4 suggests that the inverse relation between ECBQ Negative Affectivity and CBQ Effortful Control is largely explained by predictions from components of toddler Negativity to later Attention Focusing and Inhibitory Control.

DISCUSSION

The results of our investigation demonstrate the two primary advantages to the fine-grained approach to studying temperament. The robust longitudinal relations between factor scores supported the construct validity of the hierarchical structure of temperament previously revealed through factor analyses of the IBQ-R, ECBQ, and CBQ. In addition, the degree of specificity afforded by measurement of multiple distinct traits allowed increased nuance in our understanding of homotypic and heterotypic continuity.

Across adjacent measurement periods, convergent validity was demonstrated for all discrete scales considered conceptually consistent. The discriminant validity of the scales was also supported in these analyses, particularly between toddlerhood and childhood. Between the infancy and toddler assessments, 5 of 11 scales were most highly correlated with their equivalent scales, and this was true for all 14 scales found on both the ECBQ and CBQ. Multiple significant correlations were also obtained between scales from analogous factors, suggesting systemic continuity of underlying constitutional factors or common developmental factors influencing related behaviours across varied contexts.

Less stability was observed between fine-grained traits across longer developmental periods. Although Activity Level, Approach/Positive Anticipation, Smiling and Laughter, Distress to Limitations/Frustration, Sadness, and Perceptual Sensitivity were significantly stable from infancy to preschool age, even these traits demonstrated only modest continuity. This is consistent with the results of prior studies of temperament stability (see Lemery, Goldsmith, Klinnert, & Mrazek, 1999) and with the general observation that greater stability is expected over brief periods, in comparison to longer time spans. Both the elicitors and the behavioural expressions of emotions and attentional processes undergo dramatic shifts between infancy and the preschool period (Camras, 1994; Posner & Rothbart, 2007; Ruff & Rothbart, 1996), creating difficulties for consistency in measurement of fine-grained traits over this span. The emergence of new effortful capabilities is also likely to affect stability, as regulatory processes work to moderate initially reactive tendencies (Nigg, 2006; Posner & Rothbart, 2007; Rothbart & Bates, 2006).

Differential patterns of stability of Surgency, Negative Affectivity, and Effortful Control may reflect dissimilarity in developmental processes. Relations between infant and preschool Negative Affectivity were not completely mediated by toddler Negative Affectivity, possibly suggesting early consolidation of an organized system in which apparent instability across time is largely due to imperfect measurement at each age (Lemery *et al*., 1999). In contrast, infant to preschool connections for Surgency were fully mediated by toddler Surgency, perhaps indicating progressive change rather than constancy of a relatively static trait. Given the substantial development during early childhood of systems that underlie behavioural and attentional control (Posner & Rothbart, 2007; Rothbart & Rueda, 2005), it is perhaps not surprising that direct continuity was not evident at either the fine-grained or the aggregated level between infant ORC and preschool Effortful Control. It is, however, important to note the

DOI: 10.1002/icd

transactional stability of this trait: toddler Effortful Control was both predictive of preschool Effortful Control and predicted by infant ORC.

Within-factor heterotypic continuity of certain fine-grained attributes appeared to play particularly salient roles in longitudinal stability of the factors. High-Intensity Pleasure, Activity Level, and Impulsivity were longitudinally linked to several other components of Surgency, suggesting they form a 'core' of sorts, upon which other manifestations of Surgency are superimposed. These findings are reminiscent of those obtained by Komsi *et al*. (2006), in which infant Activity Level was the primary contributor to continuity between a Positive Affectivity 'superfactor' and later Surgency.

Indicating distinctions within the broad Negative Affectivity factor, Sadness, Frustration, and (reversed) Falling Reactivity were largely coherent with one another over time, but demonstrated no longitudinal relations with Fear or Shyness, consistent with neural models such as that proposed by Gray (1991), which places fear responses under the domain of a Behavioural Inhibition System (BIS) that is somewhat distinct from a Fight/Flight (F/F) system underlying irritability. These differences are also in keeping with differences between externalizing and internalizing in problem behaviour (the only exception being inclusion of sadness in the frustration rather than the fear cluster). One might speculate that general predispositions for negativity may be associated with amygdala functioning, implicated in both fear and anger, with individual differences in BIS and F/F circuitry contributing to specificity in emotional reactivity (see also Rothbart & Posner, 2006).

The fine-grained trait of Low-Intensity Pleasure appeared to play a particularly important role in the transactional stability of the regulatory factor. Because research regarding regulation of behaviour and attention has seldom considered motivational processes, this finding may be useful for directing future investigation. The ability to gain pleasure from low-intensity activities such as examining books, watching educational television, and playing cognitively demanding board games may allow children opportunities for intellectual growth. Exploration of Low-Intensity Pleasure may enhance prediction of later intelligence, which has previously been linked to less emotion-based aspects of Effortful Control (Fabes, Martin, Hanish, Anders, & Madden-Derdich, 2003).

A particularly intriguing pattern of factor-level heterotypic continuity indicated that high levels of Surgency during infancy were associated with high Effortful Control in toddlers. A basic predisposition to positive affect may partially explain these results, in that the strongest correlations were obtained between infant High-Intensity Pleasure and Smiling and Laughter and toddler Low-Intensity Pleasure. This explanation was invoked by Komsi *et al*. (2006) to explain their findings of infant Smiling and Laughter as a predictor of preschool dimensions of Effortful Control. Alternatively, high Surgency may allow a child to more frequently encounter experiences that are cognitively and affectively challenging, leading to greater opportunities to develop willful control of their behaviour and attention. A related possibility is that infants who display more positive emotionality elicit guidance from parents that may facilitate the development of attentional skills underpinning Effortful Control.

The direction of the relationship between Surgency and Effortful Control shifted in the transition from toddlerhood to the preschool period, with high Surgency between 18 and 36 months predicting low Effortful Control between 3 and 5 years. This pattern mirrors the within-age pattern of relations between these dimensions, but is nonetheless surprising due to the conceptual dissimilarity of Surgency and Effortful Control. The negative relation appears to de-

DOI: 10.1002/icd

monstrate a struggle between basic reactive and socially imposed dimensions of motivation. The more reactive systems underlying orientation and approach to reward are often at odds with later-emerging capacities to focus attention and control behaviour, and with the desire to meet societal expectations that one should do so. This would fall within the accelerator-brakes metaphor proposed by Derryberry and Rothbart (1997). Toddler Activity Level was particularly predictive of later deficits in Attentional and Inhibitory Control, reminiscent of the inverse relations between the two dimensions in the symptoms associated with attention deficit and hyperactivity in the clinical literature (e.g. Nigg, 2001).

A final finding regarding heterotypic continuity concerns inverse relations between Negative Affect and Effortful Control. High levels of several aspects of ECBQ Negative Affect were predictive of low levels of CBQ Attention Focusing and Inhibitory Control. Similar findings have been reported from childhood to adulthood (Rothbart & Sheese, 2006). The links between NA and EFF may offer support to the proposal made by Calkins (1994), Calkins and Degnan (2006) and supported during infancy by Stifter and Spinrad (2002) that temperamental negative reactivity may compromise the effective development of attentional and inhibitory mechanisms. Our results also suggest that early deficits in regulation may contribute to later negativity, insofar as high infant Orienting Regulatory Capacity forecast low preschool Negative Affect, suggesting early-emerging tendencies to focus attention, enjoy calm activities, and benefit from maternal soothing may contribute to later regulation of negative emotions.

A primary limitation of the current investigation concerns the reliance on information acquired from a single parent. Kagan (1994) has been critical of questionnaire methodology, contending that parental characteristics including personality, differing knowledge of child behaviour, and inconsistent interpretation of items are sources of bias. Potentially compounding these biases the data used to assess continuity were used by Gartstein and Rothbart (2003) and Putnam *et al*. (2006) to create the infant and toddler instruments. Because the psychometric characteristics of the scales were optimized for these samples, the stability coefficients obtained may have been overestimated. Our concerns regarding rater subjectivity were at least partially addressed in the IBQ-R and ECBQ, through utilization of items that ask only about recently occurring events, and inquiring about concrete child behaviours, rather than asking parents to make abstract or comparative judgments (Rothbart & Goldsmith, 1985). Also easing our concern, interparent agreement has been established for the IBQ-R, and ECBQ, (Gartstein & Rothbart, 2003; Putnam *et al*., 2006; Rothbart *et al*., 2001). In addition, although several studies have failed to reveal agreement between parent reports and observed behaviours (see Stifter, Willoughby, Towe-Goodman, & The Family Life Project Key Investigators 2008), others have demonstrated convergent validity between laboratory measures and the IBQ (Rothbart, 1981) and Toddler Behaviour Assessment Questionnaire (TBAQ; Goldsmith, 1996) on which the IBQ-R and ECBQ are based (Goldsmith & Rothbart, 1991), suggesting a substantial objective component to parent ratings.

An important future direction concerns the development of observational tasks to tap the constructs contained in our fine-grained measures that are not incorporated in existing, standardized laboratory batteries (e.g. Buss & Goldsmith, 2000; Goldsmith & Rothbart, 1991). It should be noted, however, that observational measures are also prone to error, including problems related to characteristics of the rater or experimenter, effects of the measure on child behaviours, and interactions between experimenter and measure characteristics (Rothbart & Goldsmith, 1985). Logistical problems also limit the scope of

DOI: 10.1002/icd

laboratory assessments, and questionnaires may be the only realistic route to assessing the wide range of traits represented in the current study.

Our study is also limited in terms of sample characteristics. Although the sample was economically diverse, it reflected the primarily Caucasian population of the Pacific Northwest. Because the psychometric properties of temperament scales may be inconsistent across samples differing in ethnicity and sociometric status (e.g. Putnam & Rothbart, 2006), subsequent studies of high-risk populations such as those carried out by Victor, Rothbart, and Baker (2006) for combined personality and temperament scales would be informative. A second limitation of our sample concerns the substantial levels of attrition. Nearly half of the families participating during the infancy period had been lost by early childhood. Although the retained and discontinuing families did not differ significantly in terms of socio-economic status, family income, parent age, or parent education, retained children were sadder and less soothable as infants. Other, unmeasured, differences may have existed between these groups of families and the generalizability of the current findings is unclear.

Our findings point to new directions, both methodological and substantive, in temperament research. Methodologically, it may be worthwhile to consider further revisions to capitalize on the longitudinal associations revealed herein. One possibility involves expanding the existing questionnaires. Due to the time course of their design (the IBQ was created first, the CBQ was created next, followed by the IBQ-R, then the ECBQ), the content of the measures reflected in part the authors' evolving thoughts regarding temperament and development. Currently, only the ECBQ contains Motor Activation and Attention Shifting scales. Similarly, recognition of an affiliation dimension that is subsumed under Effortful Control in the IBQ-R and ECBQ, but emerges as a separate factor in adolescence (Ellis & Rothbart, 2001; Putnam *et al*., 2001) and demonstrates some similarity with the personality construct of agreeableness (Evans & Rothbart, 2007) indicates the promise of adding Sociability and Cuddliness scales to the CBQ.

Conversely, these questionnaires could be shortened to reflect the nature of heterotypic within-factor continuity discerned in this research. For instance, because High-Intensity Pleasure, Activity Level, and Impulsivity appear to form a 'core' of Surgency that is longitudinally stable, instrument revisions may include trimming of more peripheral traits such as Smiling and Laughter or Positive Anticipation. Short and Very Short Forms of the CBQ have been recently developed, with the short version retaining 6- to 8-item versions of all scales from the standard CBQ, and the very short version containing three broad-based 12-item scales covering the content of the overarching factors (Putnam & Rothbart, 2006), and similar treatment of the IBQ-R and ECBQ is recommended. An additional direction for trimming involves removal of items that conceptually overlap with behaviour problem dimensions, as recommended by Lemery, Essex, and Smider (2002). It is possible to envision multiple versions of the measures that are tailored to the needs of individual researchers.

Substantively, significant stability across adjacent measurement periods was demonstrated for all conceptually consistent fine-grained traits and factors. The demonstrated developmental coherence of these factors suggests that the IBQ-R, ECBQ, and CBQ can be meaningfully combined in future studies utilizing growth-curve analyses to examine influences maintaining and lawfully altering longitudinal stability.

A particularly exciting new theoretical direction is indicated by the novel findings of longitudinal heterotypic relations between Surgency and Effortful Control. Effortful Control has been strongly implicated in the aetiology

DOI: 10.1002/icd

of behaviour problems (Calkins & Degnan, 2006; Nigg, 2006; Rothbart & Bates, 2006), but early indicators of this trait are largely unknown. Additional research in the laboratory and the home, as well as through caregiver-report, to reveal the mechanisms through which infant tendencies towards positive affect in multiple contexts are transformed into capabilities for attentional control in early childhood, is warranted. Alternatively, high Activity Level in toddlerhood appears to be a risk factor for later self-regulation, possibly indicating routes for intervention to prevent externalizing behaviour problems.

We believe that the research described here is consistent with the aims of this collection, representing an important new approach illuminating the diversity of behaviours that can be considered 'temperament'. We further feel that continuing expansion and refinement of the constructs and issues addressed in this report will generate insight into developmental processes of reactivity and regulation.

ACKNOWLEDGEMENTS

This research was supported by funds allocated to Samuel Putnam and Maria Gartstein from NIMH (US) grant 5 T32 MH1893, awarded to University of Oregon, and from a new faculty award from Bowdoin College, awarded to Samuel Putnam. Preliminary findings were presented to the International Society for Infant Studies in May 2002.

REFERENCES

Bates, J. E., Freeland, C. A., & Lounsbury, M. L. (1979). Measurement of infant difficultness. *Child Development*, *50*, 794–803.

Bell, R. Q. (1974). Contributions of human infants to caregiving and social interaction. In M. Lewis & L. A. Rosenblum (Eds.), *The effect of the infant on its caregiver* (pp. 1–19). New York: Wiley.

Buss, A. H., & Plomin, R. (1975). *A temperament theory of personality development*. New York: Wiley.

Buss, K. A., & Goldsmith, H. H. (2000). *Manual and Normative Data for the Laboratory Temperament Assessment Battery—Toddler Version*. Psychology Department Technical Report, University of Wisconsin, Madison.

Calkins, S. D. (1994). Origins and outcomes of individual differences in emotional regulation. *Monographs of the Society for Research in Child Development*, *59*(2–3), 250–283.

Calkins, S. D., & Degnan, K. A. (2006). Temperament in early development. In R. T. Ammerman (Ed.), *Comprehensive handbook of personality and psychopathology*, Vol. 3 (pp. 64–84). Hoboken, NJ: Wiley.

Camras, L. A. (1994). Two aspects of emotional development: Expression and elicitation. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 347–351). New York: Oxford University Press.

Caspi, A. 1998. Personality development across the life course. In W. Damon (Series Ed.) & N. Eiesenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (5th ed)(pp. 311–388). New York: Wiley.

Derryberry, D., & Rothbart, M. K. (1997). Reactive and effortful processes in the organization of temperament. *Development and Psychopathology*, *9*, 633–652.

Ellis, L. K., & Rothbart, M. K. (2001). Revision of the Early Adolescent Temperament Questionnaire. Poster presented at the *2001 Biennial Meeting of the Society for Research in Child Development*, Minneapolis, MN.

Else-Quest, N. M., Hyde, J. S., Goldsmith, H. H., & Van Hulle, C. A. (2006). Gender differences in temperament: A meta-analysis. *Psychological Bulletin*, *132*, 33–72.

DOI: 10.1002/icd

Evans, D. E., & Rothbart, M. K. (2007). Developing a model for adult temperament. *Journal of Research in Personality, 41*, 868–888.
Fabes, R. A., Martin, C. L., Hanish, L. D., Anders, M. C., & Madden-Derdich, D. A. (2003). Early school competence: The roles of sex-segregated play and effortful control. *Developmental Psychology, 59*, 848–858.
Fox, N. A., & Henderson, H. A. (1999). Does infancy matter? Predicting social behavior from infant temperament. *Infant Behavior and Development*, *22*, 445–455.
Gartstein, M. A., & Rothbart, M. K. (2003). Studying infant temperament via the revised infant behavior questionnaire. *Infant Behavior and Development*, *26*, 64–86.
Goldsmith, H. H. (1996). Studying temperament via construction of the Toddler Behavior Questionnaire. *Child Development*, *67*, 218–235.
Goldsmith, H. H., & Rothbart, M. K. (1991). Contemporary instruments for assessing early temperament by questionnaire and in the laboratory. In L. Strelau & A. Angleitner (Eds.), *Explorations in temperament: International perspectives on theory and measurement* (pp. 249–272). New York: Plenum.
Gray, J. A. (1991). The neuropsychology of temperament. In J. Strelau & A. Angleitner (Eds.), *Explorations in temperament: International perspectives on theory and measurement* (pp. 105–128). New York: Plenum Press.
Izard, C. E. (1977). *Human emotions*. New York: Plenum.
Kagan, J. (1969). The three faces of continuity in human development. In D. A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 26–74). Cambridge, MA: Harvard University Press.
Kagan, J. (1994). *Galen's prophecy: Temperament in human nature*. New York: Basic Books.
Kagan, J., Snidman, N., & Arcus, D. (1998). Childhood derivatives of high and low reactivity in infancy. *Child Development*, *69*, 1483–1493.
Komsi, N., Raikkonen, K., Pesonen, A., Heinonen, K., Keskivaara, P., Jarvenpaa, A., *et al*. (2006). Continuity of temperament from infancy to middle childhood. *Infant Behavior and Development*, *29*, 494–508.
Kopp, C. B. (1982). Antecedents of self-regulation: A developmental perspective. *Developmental Psychology, 18*, 199–214.
Lemery, K. S., Essex, M. J., & Smider, N. A. (2002). Revealing the relation between temperament and behavior problem symptoms by eliminating measurement confounding: Expert ratings and factor analysis. *Child Development*, *73*, 867–882.
Lemery, K. S., Goldsmith, H. H., Klinnert, M. D., & Mrazek, D. A. (1999). Developmental models of infant and childhood temperament. *Developmental Psychology, 35*, 189–204.
McDevitt, S. C., & Carey, W. B. (1978). The measurement of temperament in 3–7 year old children. *Journal of Child Psychology and Psychiatry, 19*, 245–253.
Nigg, J. T. (2001). Is ADHD a disinhibitory disorder? *Psychological Bulletin*, *127*, 571–598.
Nigg, J. T. (2006). Temperament and developmental psychopathology. *Journal of Child Psychology and Psychiatry, 47*, 395–422.
Norusis, M. J. (1994). *SPSS professional statistics 6.1*. Chicago: SPSS, Inc.
Posner, M. I., & Rothbart, M. K. (2007). *Educating the human brain*. Washington, DC: American Psychological Association.
Putnam, S. P., Ellis, L. K., & Rothbart, M. K. (2001). The structure of temperament from infancy through adolescence. In: A. Eliasz & A. Angleitner (Eds.), *Advances/proceedings in research on temperament* (pp. 165–182). Germany: Pabst Scientist Publisher.
Putnam, S. P., Gartstein, M. A., & Rothbart, M. K. (2006). Measurement of fine-grained aspects of toddler temperament: The Early Childhood Behavior Questionnaire. *Infant Behavior and Development*, *29*(3), 386–401.
Putnam, S. P., & Rothbart, M. K. (2006). Development of short and very short forms of the Children's Behavior Questionnaire. *Journal of Personality Assessment*, *87*, 102–112.
Putnam, S. P., Rothbart, M. K., & Gartstein, M. A. (2004). Longitudinal examination of fine-grained aspects of temperament during infancy, toddlerhood, and early childhood. Poster presented at the *International Conference on Infant Studies*, Chicago, IL.
Putnam, S. P., & Stifter, C. A. (2005). Behavioral approach-inhibition in toddlers: Prediction from infancy, positive and negative affective components, and relations with behavior problems. *Child Development*, *76*(1), 212–226.

DOI: 10.1002/icd

Rothbart, M. K. (1981). Measurement of temperament in infancy. *Child Development*, *52*, 569–578.
Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. *Child Development*, *72*, 1394–1408.
Rothbart, M. K., & Bates, J. E. (2006). Temperament. In W. Damon, R. Lerner, & N. Eisenberg (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed) (pp. 99–166). New York: Wiley.
Rothbart, M. K., & Derryberry, D. (1981). Development of individual differences in temperament. In M. L. Lamb & A. L. Brown (Eds.), *Advances in developmental psychology*, (Vol. 1, pp. 37–86). New York: Guilford.
Rothbart, M. K., Derryberry, D., & Hershey, K. (2000). Stability of temperament in childhood: Laboratory infant assessment to parent report at seven years. In V. J. Molfese & D. L. Molfese (Eds.), *Temperament and personality development across the life span* (pp. 85–119). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
Rothbart, M. K., & Goldsmith, H. H. (1985). Three approaches to the study of infant temperament. *Developmental Review*, *5*, 237–260.
Rothbart, M. K., & Posner, M. I. (2006). Temperament, attention, and developmental psychopathology. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology. Volume 2 Developmental neuroscience* (2nd ed) (pp. 465–501). Hoboken, NJ: Wiley.
Rothbart, M. K., & Rueda, M. R. (2005). The development of effortful control. In U. Mayr, E. Awh, & S. W. Keele (Eds.), *Developing individuality in the human brain: A tribute to Michael I. Posner* (pp. 167–188). Washington, DC: American Psychological Association.
Rothbart, M. K., & Sheese, B. (2006). Temperament and emotion regulation. In J. J. Gross (Ed.), *Handbook of emotion regulation*. New York: Guilford Press.
Ruff, H. A., & Rothbart, M. K. (1996). *Attention in early development: Themes and variations*. New York: Oxford University Press.
Sankoh, A. J., Huque, M. F., & Dubey, S. D. (1997). Some comments on frequently used multiple endpoint adjustment methods in clinical trials. *Statistics in Medicine*, *16*, 2529–2542.
Sanson, A., Prior, M., Garino, E., & Oberklaid, F. (1987). The structure of infant temperament: Factor analysis of the Revised Infant Temperament Questionnaire. *Infant Behavior and Development*, *10*, 97–104.
Stevens, G., & Featherman, D. L. (1981). A revised socioeconomic index of occupational status. *Social Science Research*, *10*, 365–395.
Stifter, C. A., & Spinrad, T. L. (2002). The effect of excessive crying on the development of emotion regulation. *Infancy*, *3*, 133–152.
Stifer, C. A., Willoughby, M. T., Towe-Goodman, N., & The Family Life Project Key Investigators. (2008). Agree or agree to disagree? Assessing the convergence between parents and observers on infant temperament. *Infant and Child Development*.
Thomas, A., Chess, S., & Birch, H. G. (1968). *Temperament and behavior disorders in children*. New York: New York University Press.
Uitenbroek, D. G. (1997). *SISA Binomial*. Southampton: D.G. Uitenbroek. Retrieved July 11, 2006, from the World Wide Web: http://home.clara.net/sisa/bonhlp.htm.
Victor, J. B., Rothbart, M. K., & Baker, S. R. (2006). Temperamental components of the developing personality. Unpublished manuscript.

DOI: 10.1002/icd