Available online at www.sciencedirect.com
SCIENCE DIRECT
Infant Behavior & Development 29 (2006) 386–401
Infant Behavior & Development

Measurement of fine-grained aspects of toddler temperament: The Early Childhood Behavior Questionnaire

Samuel P. Putnam [a,*], Maria A. Gartstein [b,1], Mary K. Rothbart [b]

[a] *Department of Psychology, Bowdoin College, 6900 College Station, Brunswick, ME 04011, United States*
[b] *Department of Psychology, University of Oregon, United States*

Received 27 June 2005; received in revised form 20 January 2006; accepted 26 January 2006

Abstract

This article describes the development, reliability, and factor structure of a finely differentiated (18 dimensions) parent-report measure of temperament in 1.5- to 3-year-old children, using a cross-sectional sample ($N = 317$) and a longitudinal sample of primary ($N = 104$) and secondary ($N = 61$) caregivers. Adequate internal consistency was demonstrated for all scales and moderate inter-rater reliability was evident for most scales. Longitudinal stability correlations were primarily large over 6- and 12-month spans and moderate to large from 18 to 36 months. Factor analysis revealed a three-factor structure of Surgency/Extraversion, Negative Affectivity, and Effortful Control. In both samples and for both primary and secondary caregivers, older children received higher scores for Attention Focusing, Discomfort, Inhibitory Control, and Positive Anticipation. Primary caregivers rated females higher in Fear, and lower in High-intensity Pleasure, than males; secondary caregivers rated females higher than males in several aspects of Effortful Control.

Keywords: Temperament; Assessment; Toddlers

Increases in the number of empirical studies of temperament in recent years (Rothbart & Bates, 1998, 2006) have been matched by an expanded and more finely defined list of dimensions considered within the temperament realm. Although fine-grained instruments have been developed to assess temperament in infants (Gartstein & Rothbart, 2003) and older children (Rothbart, Ahadi, Hershey, & Fisher, 2001), an analogous instrument appropriate for use with children between the ages of 1 and 3 has not been made available. The current study describes the psychometric characteristics, factor structure, and demographic correlates of an instrument designed to fill this gap: the Early Childhood Behavior Questionnaire (ECBQ).

The ECBQ was originally designed to supplement the Toddler Behavior Assessment Questionnaire (TBAQ; Goldsmith, 1996), a widely used parent-report temperament questionnaire for young children. The TBAQ includes 108 items that address five aspects of temperament: Activity Level, Pleasure, Social Fearfulness, Anger Proneness, and Interest/Persistence. Goldsmith (1996) documented internal consistency and inter-rater reliability of the instrument, as well as convergence with other temperament measures. Subsequent studies (e.g., Eiden, Edwards, & Leonard, 2004;

* Corresponding author. Tel.: +1 207 725 3152; fax: +1 207 725 3892.
E-mail address: sputnam@Bowdoin.edu (S.P. Putnam).
[1] Present address: Department of Psychology, Washington State University, United States.

0163-6383/$ – see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.infbeh.2006.01.004

Kochanska & Knaack, 2003; Lemery, Goldsmith, Klinnert, & Mrazek, 1999) have provided support for the construct validity of the instrument. The value of this measure is further indicated by successful translations to Japanese (Kusanagi, Chen, & Hoshi, 2000), Spanish (Salinas, Montesinos, & Carnicero, 1999), and Dutch (Van Bakel & Riksen-Walraven, 2004).

Several aspects of temperament assessed by the new instrument, but not the TBAQ, reflect differences between our theoretical approach and the approach that guided the creation of the TBAQ. Whereas Goldsmith (1996) regarded temperament solely in terms of individual differences in emotionality during early development, the ECBQ is based on a definition of temperament that includes reactive processes involving not only emotion, but also motor and sensory systems, as well as an emphasis on self-regulatory processes that modulate reactivity (Rothbart & Derryberry, 1981; Rothbart et al., 2001). Within the domain of emotion, the ECBQ is broader in content than the TBAQ. Rather than a single Pleasure scale, for example, the new measure separately assesses pleasure related to low- and high-intensity activities, and also includes a scale concerning pleasure about upcoming activities. Differences also exist in the individual items within the scales shared by the two instruments. To ease time demands on subjects, the ECBQ uses a stem-and-leaf format, in which a single context (e.g., "When playing outdoors, how often did your child") is followed by multiple responses (e.g., "like making lots of noise", "enjoy sitting quietly in the sunshine", "want to climb to high places"). In addition, several scales include entirely new items not included on the earlier measure. Due to these substantial differences between the two measures, the name Early Childhood Behavior Questionnaire (ECBQ) was chosen, rather than TBAQ-Revised.

1. Temperament constructs assessed in the new instrument

The 18 scales included in the ECBQ are predominantly "downward extensions" of dimensions contained on the Children's Behavior Questionnaire (CBQ; Rothbart et al., 2001) and "upward extensions" from the Infant Behavior Questionnaire-Revised (IBQ-R; Gartstein and Rothbart, 2003). Ten scales are found in similar forms on all three instruments: Activity Level, Attention Focusing (labeled Duration of Orienting on IBQ-R and including attention shifting capability on the CBQ), Fear (including startle and reactions to both social and non-social stimuli on the IBQ-R), Frustration, High- and Low-intensity Pleasure, Perceptual Sensitivity, Positive Anticipation (measured as rate of approach in the IBQ-R), Sadness, and Soothability. One scale, Affiliation/Cuddliness is found on the IBQ-R, but not the CBQ. Four scales included in the new instrument, Discomfort, Impulsivity, Inhibitory Control, and Shyness, are included on the CBQ, but not the IBQ-R. Finally, three scales on the ECBQ, Attention Shifting, Motor Activation, and Sociability, are not included on either the CBQ or IBQ-R. See Putnam, Ellis, and Rothbart (2001) for further information concerning decisions regarding inclusion of scales in some instruments, but not others.

2. Temperament structure

A fine-grained approach to temperament, in addition to allowing greater specificity in predicting and assessing relations between temperament and other constructs, contributes to an understanding of temperament through investigation of hierarchical relations among traits. Factor analysis of the IBQ-R and CBQ has consistently yielded a three-factor structure (Gartstein & Rothbart, 2003; Kochanska, DeVet, Goldman, Murray, & Putnam, 1994; Rothbart et al., 2001). In both infants and older children, a Surgency factor is marked by high loadings for Activity Level and High-intensity Pleasure. Impulsivity and Shyness (loading negatively) further define this factor for the CBQ. Positive Anticipation holds its primary loading on Surgency for the IBQ-R, and a high secondary loading on this factor for the CBQ. A second factor, Negative Affectivity, is marked by primary positive loadings for Frustration, Sadness, and Fear for both the IBQ-R and CBQ, and also by Discomfort and Positive Anticipation on the CBQ. Negative primary loadings on Negative Affectivity are obtained for Falling Reactivity for the IBQ-R and Soothability for the CBQ. Finally, a third factor, labeled Orienting/Regulation in infants and Effortful Control in children, is distinguished by primary loadings for Low-intensity Pleasure and Duration of Orienting/Attention Control in both measures. Soothability and Cuddliness further define the factor in the IBQ-R, whereas Inhibitory Control, Smiling and Laughter, and Perceptual Sensitivity complete the CBQ factor.

Based on the findings regarding temperament structure of fine-grained temperament in infants and older children, it was expected that factor analyses would reveal three factors, with High-intensity Pleasure, Activity Level,

reversed Shyness, and Impulsivity anchoring Surgency; Frustration, Sadness, Fear and Discomfort loading primarily on Negative Affectivity; and Attention Focusing, Attention Shifting, Inhibitory Control, Cuddliness, and Low-intensity Pleasure defining Effortful Control. Because factor affiliation was inconsistent between the IBQ-R and CBQ for Soothability, Perceptual Sensitivity, and Positive Anticipation, no predictions were made regarding these scales and their roles in defining the ECBQ factor structure. Further, a priori hypotheses could not be generated for Sociability and Motor Activation, because these scales have not been included on either the IBQ-R or CBQ.

3. Development of temperament during toddlerhood

Temperament research has typically emphasized rank-order stability of individual differences. Recent perspectives (e.g., Rothbart, 1989, 1994), however, suggest that the neural, motor, and cognitive factors underlying reactivity and regulation are not fully developed at birth, but that maturation leads to changes, or mean-level instability, in the expression of temperament dimensions. Across TBAQ scales, Goldsmith (1996) noted a slight tendency for scale scores to increase with age, with the exception of Social Fearfulness, which decreases slightly during toddlerhood. In reference to these findings, we expected to document increases in aspects of Surgency and non-fearful Negative Affectivity. The most profound changes during this period, however, were expected for scales associated with Effortful Control. Posner and Rothbart (2000), and Rothbart and Rueda (2005) contend that rapid development of frontal neural systems, particularly the anterior cingulate, during early childhood may underlie trends toward greater control of attention and behavior. Consistent with this reasoning, the ability to exercise inhibitory control in delaying gratification increases from 18 to 30 months (Vaughn, Kopp, & Krakow, 1984), and focused attention during free play increases between 2.5 and 4.5 years (Ruff & Lawson, 1990).

3.1. Gender differences in temperament

Although Goldsmith (1996) found minor and inconsistent gender effects for toddler temperament, other studies have revealed gender differences in infancy and childhood. Activity level may be the trait for which sex differences are observed earliest, as male fetuses have been reported to exhibit more frequent leg movements than female fetuses (Almli, Ball, & Wheeler, 2001). A recent meta-analysis (Campbell & Eaton, 1999) demonstrates consistent findings, based on multiple measurement strategies, of higher activity level in male than female infants. Differences in approach–withdrawal behavior, suggesting elevated Fear, Shyness, and withdrawal in females, and greater High-intensity Pleasure seeking in males, have also been documented during infancy and early toddlerhood (Carey & McDevitt, 1978; Gartstein & Rothbart, 2003; Martin, Wisenbaker, Baker, & Huttunen, 1997; Maziade, Broudreault, Thivierge, Caperaa, & Cote, 1984; Reznick, Gibbons, Johnson, & McDonough, 1989). These sex differences persist throughout childhood (Cote, Tremblay, Nagin, Zoccolillo, & Vitaro, 2002; Eaton & Enns, 1986). Other sex differences have been observed at older ages. Eiden et al. (2004) found girls to be higher in effortful control at 2 years of age, and Ahadi, Rothbart, and Ye (1993) found maternal ratings of Inhibitory Control, Low-Intensity Pleasure, and Perceptual Sensitivity in 6- and 7-year-old children to be higher for American girls than boys, but conversely related in a Chinese sample.

The current paper describes the development and psychometric characteristics of the ECBQ, including assessment of internal consistency, inter-rater agreement, and stability of scores from 18 to 36 months. In addition, the structure of the 18-scale instrument is investigated using factor analysis, and age- and gender-based differences in temperament are examined. It is expected that all scales will demonstrate adequate internal consistency, and that inter-rater agreement and longitudinal stability of scale scores will approximate levels achieved with other measures of toddler temperament. We also anticipate emergence of a structure containing three factors representing Surgency/Extraversion, Negative Affectivity, and Effortful Control. It is hypothesized that higher scores with increasing child age will be obtained for most scales, with the possible exceptions of Shyness, Impulsivity, and Soothability. Finally, we predict that females will be rated more highly on Fear, Shyness, Inhibitory Control, Low-Intensity Pleasure, and Perceptual Sensitivity; and higher scores will be obtained for males on Activity Level, Positive Anticipation, and High-intensity Pleasure.

4. Method

4.1. Participants

We describe analyses carried out on two separate samples. Sample 1 was given a preliminary version of the measure and was used to make decisions regarding item retention. An attempt was made to avoid a common problem associated with questionnaire refinement: reporting only those data collected with a larger instrument from which items had been removed, a practice that tends to overestimate internal consistency (Smith, McCarthy, & Anderson, 2000). Thus, a second sample was recruited and administered the final version of the form in the present study.

The majority of Sample 1 participants were recruited between the ages of 3 and 12 months from the Eugene–Springfield, Oregon area on the basis of birth announcements in the local newspaper and participation in local "Birth to 3" courses. Parents of the 361 (181 female) infants originally recruited were predominantly Caucasian, middle class (average family income = US$ 41,798.30 [S.D. = 19,154.50) and employed in service-oriented professions (Revised Duncan Sociometric Index [Stevens & Featherman, 1981] mean of 35). Primary caregivers had an average age of 31 (S.D. = 5.30), had completed an average of 14.51 (S.D. = 2.40) years of school, and 96% were married (see Gartstein & Rothbart, 2003 for details regarding the composition of this sample). From this group of 361, primary caregivers of 252 (126 female) children completed the toddler measure. Of the 109 participants not participating, 27 declined participation when contacted by phone, 55 agreed to participate but did not return questionnaires following multiple reminders by phone and mail, and forwarding information could not be obtained for 24. A comparison of those who did not participate with those who did 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 an additional 77 children were contacted through birth announcements, with 68 (39 female) agreeing to participate. Demographic information was not gathered from these 68 families, but the similarity in recruitment strategy allows some degree of confidence for our inference that they represent the same population as the original sample. The total sample of 317 children (164 female) was roughly equally distributed across three age groups: 18–22 months ($n = 103$; 54 female), 22–26 months ($n = 110$; 55 female), and 27–32 months ($n = 104$; 55 female).

Sample 2 participants were recruited from mid-coast Maine on the basis of birth announcements, flyers placed in local daycares and pediatricians' offices, and advertisements in the local newspaper. One-hundred and eighteen families contacted the research coordinator, with 104 (88%) agreeing to participate. Parents were predominantly Caucasian (95% of mothers; 90% of fathers) and married or cohabitating (93%) at the time of recruitment. Mothers' ages ranged from 19 to 42 ($M = 31.8$, S.D. = 5.66), fathers' ages ranged from 21 to 48 ($M = 33.50$, S.D. = 6.18), and annual household income ranged from US$ 5000 to 250,000 ($M = 59{,}794$, S.D. = 44,388; data missing for 11 families). Primary caregivers had an average of 15.12 (S.D. = 2.26) years of education and secondary caregivers had attended school for an average of 14.58 (S.D. = 2.56) years. Primary caregivers completed ECBQs when children were 18 ($n = 104$), 24 ($n = 99$), 30 ($n = 98$), and 36 ($n = 94$) months of age. A small number of secondary caregivers (ns = 55, 43, 51, and 48 at the four time points) also completed forms on a volunteer basis. With three exceptions, primary caregivers were mothers and secondary caregivers were fathers. In two families, these roles were reversed, and in one other, the child's grandmother acted as a second caregiver.

4.2. Procedure

4.2.1. Instrument construction

As with the development of the Infant Behavior Questionnaire (Rothbart, 1981), IBQ-R (Gartstein & Rothbart, 2003), TBAQ (Goldsmith, 1996), and CBQ (Rothbart et al., 2001), scale items describing a frequently occurring context (e.g., "When being dressed or undressed," "When playing outdoors", "When told no") and asking parents to report on the frequency of specific child behaviors (e.g., how often did your child "sit quietly and watch," "become sadly tearful") in 7-point Likert-style format ranging from "never" to "always" were rationally generated with reference to precise conceptual definitions of temperament dimensions. An iterative process was used to arrive at the items included on the final version of the instrument. In an initial effort, a sample of 138 primary caregivers of 17–31-month old children were administered an 144-item instrument containing the 65 items from a short version of the TBAQ, in addition to 79 new items. Only Goldsmith's TBAQ items were used for the Social Fear, Pleasure, Interest, Activity Level and Anger scales, and new items were added to Goldsmith's Inhibitory Control scale. The new scales included

Discomfort, Positive Anticipation, High-Intensity Pleasure, Low-Intensity Pleasure, Perceptual Sensitivity, Sadness, Attention Focusing, and Attention Shifting.

Items that lowered the internal consistency of scales, or correlated highly with scales other than those they were intended to measure, were not considered for the subsequent version of the measure. In addition, due to conceptual overlap with the new positive-affect and attention scales, the TBAQ Interest and Pleasure scales were removed from consideration. New items were generated for the remaining scales, as were items for the following additional scales: Cuddliness, (non-social) Fear, Impulsivity, Motor Activation, Soothability, and Sociability. All items were worded to fit a stem-and-leaf format. The resulting instrument, containing 18 scales and 267 items, was administered to Sample 1 of the current study. Items lowering alpha or correlating highly with other scales in these data were removed to arrive at the final 201 items and 18 scales of the ECBQ.

4.2.2. Data collection

Sample 1 parents were mailed consent forms, either a US$ 5 check or a US$ 5 gift certificate to a local toy store, the 267-item version of the ECBQ described in the preceding section, and a self-addressed stamped envelope to return the completed forms. Reminder calls and letters were used to contact parents who did not return questionnaires within three weeks. A second reminder was given to those who had not returned forms six weeks after mailing, and a third reminder was sent 1 month after this.

One month prior to the child's 18-, 24-, 30-, and 36-month birthdays, Sample 2 caregivers were mailed packets containing consent forms, the ECBQ, and other questionnaires. For the 18-month collection, parents were also sent a self-addressed stamped envelope to return the completed forms. At the later time points, parents were asked to complete the questionnaires and bring them to scheduled visits to our laboratory. A second set of forms was sent to families in which a secondary caregiver had agreed to participate. Upon receipt of the questionnaires and completion of lab visits, families were sent checks for US$ 10 at 18 months, US$ 20 at 24 and 30 months, and US$ 50 at 36 months.

5. Results

5.1. Internal consistency

Alpha coefficients were calculated separately for the three age groups in Sample 1 and the four time points in Sample 2. Of the 54 alphas calculated for Sample 1, 28 were over .80, indicating very good internal consistency (DeVellis, 1991). Only eight alphas were below .70, and only three (Impulsivity from 18 to 22 months, $\alpha = .58$, and from 22 to 26 months, $\alpha = .54$ Attention Shifting from 22 to 26 months, $\alpha = .58$) were below .60, the value considered the threshold for adequate internal consistency by DeVellis (1991).

Because internal consistency was expected to be inflated in the data set used to construct the questionnaire, it was reassuring that alpha coefficients for Sample 2 closely approximated those for Sample 1. Alpha coefficients for the 18 scales at the 4 age points can be found in Table 1. Of the 72 estimates calculated, 37 were over .80. Only five coefficients, including three at 18 months, were below .70, and only one (Impulsivity at 18 months) was below .60.

Similar levels of internal consistency were obtained from secondary caregivers' scores. Of 72 coefficients, 34 were over .80. Eleven were below .70. Only Impulsivity generated alphas below .60: αs = .59, .56, .57 at 24, 30, and 36 months, respectively.

5.2. Factor structure

Prior to examining relations among ECBQ scales, and due to significant correlations between scale scores from different assessment phases, Sample 2 data from primary caregivers were first consolidated by averaging across the four time points. To maximize *N* in Sample 1, data from all participants were included, rather than conducting separate analyses for the three groups. Following Rothbart et al. (2001), and Gartstein and Rothbart (2003), principal axis extraction was utilized, and extracted factors were obliquely rotated using the Oblimin algorithm (Norusis, 1994) to examine higher order relations between scales. Due to a small ratio of respondents to variables, factor analysis was not performed on Sample 2 secondary caregiver data.

The three-factor solutions for Samples 1 and 2 primary caregivers are presented in Table 2. One factor was similar in nature to the Negative Affectivity dimensions found in the IBQ-R and CBQ, with primary loadings for Discomfort,

Table 1
Cronbach's alphas for ECBQ scales for Sample 2 primary caregivers

Scale	Number of items	Alpha			
		18 months	24 months	30 months	36 months
Discomfort	10	.76	.84	.83	.89
Fear	11	.72	.73	.85	.81
Motor Activation	11	.75	.75	.85	.78
Sadness	12	.79	.85	.87	.84
Perceptual Sensitivity	12	.90	.82	.86	.89
Shyness	12	.78	.84	.85	.81
Soothability	9	.77	.88	.84	.84
Frustration	12	.76	.76	.83	.87
Impulsivity	10	.57	.76	.71	.70
Activity Level	12	.66	.60	.71	.71
High-Intensity Pleasure	12	.82	.84	.83	.88
Sociability	8	.88	.89	.85	.86
Positive Anticipation	11	.85	.81	.85	.82
Inhibitory Control	12	.86	.89	.89	.90
Attention Shifting	12	.62	.74	.75	.73
Low-Intensity Pleasure	11	.76	.75	.67	.77
Cuddliness	12	.85	.85	.88	.87
Attention Focusing	12	.86	.81	.90	.87

Note: $N = 104$, 99, 98, and 94 at 18, 24, 30, and 36 months, respectively.

Fear, Sadness, Frustration, and Soothability (loading negatively). Motor Activation, Perceptual Sensitivity, and Shyness also loaded primarily on this factor. The Negative Affectivity factor was the third to emerge and explained 13% of the variance in Sample 1; it emerged first and explained 25% of the variance in Sample 2. The second factor appeared to represent Surgency/Extraversion, including primary loadings for Impulsivity, Activity Level, High-intensity Pleasure, Sociability, and Positive Anticipation. This factor emerged second in Sample 1 and third in Sample 2, accounting for 14% of the variance in each sample. The third factor appeared to correspond to Effortful Control, and was defined primarily by loadings of Inhibitory Control, Attention Shifting, Low-intensity Pleasure, Cuddliness, and Attention

Table 2
Factor loadings for Sample 1 (S1; $N = 317$) and Sample 2 primary caregivers (S2; $N = 104$)

Scale	Negative Affectivity		Surgency–Extraversion		Effortful Control	
	S1	S2	S1	S2	S1	S2
Discomfort	.70	.87				
Fear	.68	.76				
Motor Activation	.40	.63				
Sadness	.57	.64				
Perceptual Sensitivity	.38	.58			.37	
Shyness	.47	.55	−.37			
Soothability	−.42	−.51			.33	.46
Frustration	.58	.50			−.32	−.44
Impulsivity			.52	.70		
Activity Level			.64	.70		−.37
High-intensity Pleasure			.57	.67		
Sociability			.49	.58		
Positive Anticipation			.51	.48		
Inhibitory Control					.66	.73
Attention Shifting					.70	.72
Low-intensity Pleasure					.64	.61
Cuddliness					.51	.44
Attention Focusing					.51	.43

Note: Principle axis factoring/Oblimin oblique rotation; loadings < .30 not included.

Table 3
Cross-rater agreement for ECBQ scales and factor summary scores

Scale	18 months	24 months	30 months	36 months	Combined across age
Discomfort	.61**	.58**	.36**	.43**	.55**
Fear	.53**	.37*	.47**	.39**	.53**
Motor Activation	.25#	.09	.13	.23	.40**
Sadness	.36**	.36*	.36*	.45**	.42**
Perceptual Sensitivity	.20	.36*	.23	.28#	.31*
Shyness	.42**	.55**	.49**	.44**	.57**
Soothability	.30**	.40*	.22	.22	.30*
Frustration	.57**	.59**	.42**	.20	.53**
Impulsivity	.19	.39*	.30*	.27#	.36**
Activity Level	.45**	.28#	.54**	.35*	.47**
High-Intensity Pleasure	.23#	.27#	.48**	.31*	.33**
Sociability	.12	.43**	.12	−.04	.32*
Positive Anticipation	.25#	.27#	.27#	.11	.36**
Inhibitory Control	.50**	.40*	.42**	.55**	.57**
Attention Shifting	.21	.04	.48**	−.01	.25#
Low-Intensity Pleasure	.14	−.15	.02	.09	.09
Cuddliness	.29*	.37*	.41**	.33*	.36**
Attention Focusing	.18	.42*	.23	.37*	.24#
Factor					
Negative Affectivity	.48**	.42*	.30*	.27#	.45**
Surgency/Extraversion	.21	.33*	.29*	.24	.31*
Effortful Control	.33*	.27#	.44**	.25#	.36**

Note: $N = 55$, 43, 51, 47, and 61 at 18, 24, 30, and 36 months, and combined across age, respectively.
* $p < .05$.
** $p < .01$.
# $p < .10$.

Focusing. This factor was the first to emerge from the Sample 1 data and the second to emerge for Sample 2, explaining 21% and 17% of the variance, respectively. Correlations between the factors were small in size: for Samples 1 and 2 primary caregivers, respectively, rs = .06 and .09 for Negative Affectivity and Surgency/Extraversion; corresponding rs = −.14, −.10 for Negative Affectivity and Effortful Control; and rs = −.03, −.07 for Surgency/Extraversion and Effortful Control.

On the basis of these analyses, summary scores for the broad factors were created by averaging scores of primarily loading scales (e.g., the Negative Affectivity factor is the average of the Discomfort, Fear, Sadness, Frustration, reversed Soothability, Motor Activation, Perceptual Sensitivity, and Shyness scales). When these summary scores were correlated with one another, using Sample 1, Sample 2 primary caregivers (averaged over all time points), and Sample 2 secondary caregivers (averaged over all time points), respectively, Negative Affect and Surgency were not significantly related to one another, rs (317, 104, 61) = −.04, .14, −.01, all n.s.. Negative Affect and Surgency were similarly uncorrelated, rs (317, 104, 61) = −.07, −.15, −.10, all n.s.. For all three data sources, a negative correlation was obtained between Effortful Control and Negative Affectivity, although this relationship only reached marginal levels of significance among Sample 2 primary caregivers, rs (317, 104, 61) = −.25, −.27, −.32, ps<caregivers, rs < .01, <.10, <.01.

5.3. Inter-rater reliability

Agreement between primary and secondary caregivers was assessed for Sample 2 families for which secondary caregivers provided data (ns = 55, 43, 51, and 48 at 18, 24, 30, and 36 months of age, respectively). Correlation coefficients for the 18 scales and the three factor summary scores at the four time points, and when ratings were averaged over the four times, are shown in Table 3. The most consistent agreement was found for scales assessing discrete aspects of Negative Affectivity: of the 25 correlations calculated for Discomfort, Shyness, Frustration, Fear, and Sadness, only one (Frustration at 36 months) was below .34. Performing most poorly in terms of inter-rater reliability

Table 4
Six-, 12-, and 18-month longitudinal stability of primary caregivers' ratings

Scale	6-month		12-month		18-month	
	18–24 months	24–30 months	30–36 months	18–30 months	24–36 months	18–36 months
Discomfort	.68	.75	.74	.58	.72	.63
Fear	.61	.63	.78	.56	.60	.55
Motor Activation	.68	.78	.76	.68	.77	.58
Sadness	.58	.73	.76	.53	.70	.54
Perceptual Sensitivity	.72	.66	.79	.54	.68	.62
Shyness	.58	.45	.62	.43	.41	.34
Soothability	.48	.56	.70	.49	.62	.50
Frustration	.50	.52	.68	.52	.59	.58
Impulsivity	.42	.74	.63	.37	.55	.35
Activity Level	.56	.63	.74	.52	.60	.49
High-Intensity Pleasure	.63	.63	.72	.48	.77	.49
Sociability	.56	.58	.44	.44	.59	.54
Positive Anticipation	.33	.59	.67	.35	.52	.07
Inhibitory Control	.51	.67	.67	.55	.69	.59
Attention Shifting	.52	.59	.55	.43	.66	.57
Low-Intensity Pleasure	.51	.55	.62	.50	.61	.47
Cuddliness	.58	.54	.59	.53	.46	.46
Attention Focusing	.46	.57	.68	.50	.43	.39
Factor						
Negative Affectivity	.76	.78	.83	.76	.79	.75
Surgency/Extraversion	.61	.73	.80	.50	.74	.45
Effortful Control	.57	.70	.75	.59	.71	.56

Note: $N = 99, 97, 93, 98, 93$, and 94 for 18–24, 24–30, 30–36, 18–30, 24–36, and 18–36 months, respectively. All correlations significant to $p < .01$, except Positive Anticipation from 18 to 36 months.

were Low-intensity Pleasure, Sociability, Motor Activation, Attention Shifting, and Positive Anticipation: only 2 of 20 time-specific correlations (Sociability at 24 months, Attention Shifting at 30 months) for these scales were statistically significant, although aggregating across time resulted in significant correlations for Sociability, Motor Activation, and Positive Anticipation.

5.4. Longitudinal stability

Stability coefficients for all 6- and 12-month spans, and from 18 to 36 months, are presented in Table 4 (primary caregivers) and Table 5 (secondary caregivers). With the exception of the 18–36 month correlation for Positive Anticipation, all correlations for primary caregivers are significant at $p < .01$, ranging from .32 to .79 over 6 months, and from .35 to .63 over 18 months. For data provided by secondary caregivers, 6-month correlations range from .30 to .85 over 6 months, and from .26 to .73 over 18 months. All correlations were statistically significant at $p < .05$, except High-Intensity Pleasure and Attention Focusing from 18 to 24 months, and Positive Anticipation and Cuddliness from 18 to 36 months, $ps < .10$.

5.5. Age and gender

Age and gender effects were assessed using a 2×3 (Gender $\times$ Age Group) ANOVA for Sample 1, and 2×4 (Gender $\times$ Age) repeated-measures ANOVAs for Sample 2. For descriptive purposes, means and standard deviations for Sample 2 data from primary caregivers are shown in Table 6. Primary caregivers' ratings were higher with age for 13 of the 18 dimensions. Ratings decreased with age for four dimensions: Motor Activation, Soothability, Impulsivity, and Activity Level. For only one scale, Cuddliness, were no age effects observed for the primary caregiver ratings. Significant age effects in the same directions were obtained from Sample 1 data for Attention Focusing, Discomfort, Inhibitory Control, and Positive Anticipation, $Fs(2, 311) = 13.39, 5.03, 3.03$, and 3.84, respectively, $ps < .05$. Significant age effects, in directions consistent with those for primary caregivers, were obtained from Sample 2 secondary caregivers for

Table 5
Six-, 12-, and 18-month longitudinal stability of secondary caregivers' ratings

Scale	6-month		12-month		18-month	
	18–24 months	24–30 months	30–36 months	18–30 months	24–36 months	18–36 months
Discomfort	.84	.67	.70	.70	.67	.73
Fear	.68	.62	.74	.43	.59	.48
Motor Activation	.63	.69	.69	.54	.66	.41
Sadness	.53	.37	.45	.28	.50	.54
Perceptual Sensitivity	.73	.39	.60	.53	.50	.61
Shyness	.51	.51	.36	.42	.64	.53
Soothability	.54	.60	.65	.64	.51	.61
Frustration	.69	.68	.67	.56	.77	.46
Impulsivity	.58	.46	.58	.60	.29	.49
Activity Level	.76	.58	.65	.45	.65	.60
High-Intensity Pleasure	.35	.48	.55	.58	.41	.30
Sociability	.55	.69	.62	.55	.59	.45
Positive Anticipation	.45	.76	.79	.55	.50	.26
Inhibitory Control	.69	.70	.83	.51	.72	.49
Attention Shifting	.58	.41	.71	.40	.44	.32
Low-Intensity Pleasure	.59	.60	.50	.51	.63	.42
Cuddliness	.59	.44	.65	.42	.66	.28
Attention Focusing	.28	.47	.63	.54	.63	.41
Factor						
Negative Affectivity	.80	.69	.77	.68	.74	.72
Surgency/Extraversion	.71	.73	.67	.70	.53	.41
Effortful Control	.75	.60	.77	.51	.63	.51

Note: $N = 39$, 39, 46, 46, 40, and 45 for 18–24, 24–30, 30–36, 18–30, 24–36, and 18–36 months, respectively. All correlations significant to $p < .05$, except Attention Focusing from 18 to 24 months, Impulsivity from 24 to 36 months, Sadness from 18 to 30 months, and Positive Anticipation and Cuddliness from 18 to 36 months.

Table 6
Descriptive statistics and age comparisons for Sample 2 primary caregivers' ratings

Scale	Means (standard deviations)				
	18 months	24 months	30 months	36 months	*F*
Discomfort	1.99 (.76)	2.26 (.90)	2.55 (1.02)	2.73 (1.10)	26.42**
Fear	2.17 (.81)	2.33 (.87)	2.52 (.96)	2.80 (.99)	18.02**
Motor Activation	2.27 (.82)	2.08 (.76)	2.12 (.90)	2.03 (.75)	3.42*
Sadness	2.45 (.86)	2.65 (.91)	2.75 (.90)	2.82 (.92)	6.70**
Perceptual Sensitivity	3.73 (1.13)	3.94 (1.04)	4.14 (1.05)	4.26 (1.19)	10.80**
Shyness	3.01 (.84)	3.20 (.91)	3.21 (.95)	3.52 (.85)	8.19**
Soothability	5.75 (.58)	5.54 (.78)	5.46 (.74)	5.34 (.76)	10.88**
Frustration	3.29 (.81)	3.56 (.77)	3.66 (.88)	3.65 (.93)	6.92**
Impulsivity	5.13 (.73)	5.02 (.75)	4.93 (.72)	4.78 (.69)	5.36**
Activity Level	4.98 (.73)	4.90 (.65)	4.94 (.72)	4.75 (.77)	3.87*
High-Intensity Pleasure	4.79 (.95)	4.95 (.92)	5.14 (.87)	5.07 (1.03)	3.81*
Sociability	5.39 (1.02)	5.57 (.92)	5.74 (.85)	5.81 (.87)	12.26**
Positive Anticipation	4.26 (1.10)	4.86 (.93)	5.15 (.81)	5.17 (.77)	18.64**
Inhibitory Control	3.74 (.92)	3.86 (.97)	4.08 (.94)	4.20 (.99)	11.04**
Attention Shifting	4.47 (.67)	4.59 (.68)	4.72 (.65)	4.79 (.62)	10.43**
Low-Intensity Pleasure	4.92 (.71)	4.95 (.76)	4.88 (.70)	4.89 (.71)	.19
Cuddliness	4.88 (.80)	4.94 (.78)	5.02 (.79)	5.21 (.82)	5.57**
Attention Focusing	4.08 (.85)	4.38 (.71)	4.61 (.82)	4.76 (.79)	15.67**

Note: Age × Sex repeated-measures ANOVA, d.f. = 3, 87. Values in table reflect descriptive statistics for the entire sample, $N = 104$, 99, 98, 94 at 18, 24, 30, and 36 months, respectively.

Discomfort, Fear, Soothability, Perceptual Sensitivity, Shyness, Impulsivity, Positive Anticipation, Inhibitory Control, Attention Shifting, and Attention Focusing, Fs(3, 31) = 9.20, 4.34, 3.30, 3.03, 3.10, 4.91, 5.40, 4.65, 8.08, and 14.23, respectively, ps < .05.

Age-related increases in scores on the factor summary scores were also apparent. For Sample 2 primary caregivers, Fs(3, 88) = 33.35, 7.02, and 18.88, ps < .01, for Negative Affectivity, Surgency, and Effortful Control, respectively. For Sample 2 secondary caregivers, age effects were significant for Negative Affectivity, $F(3, 31) = 5.10$, $p < .01$, and Effortful Control, $F(3, 28) = 6.40$, $p < .01$. For Sample 1 data, age effects were only significant for Effortful Control, $F(2, 311) = 6.55$, $p < .01$.

Sample 1 caregivers rated females higher than males in Fear, $F(1, 311) = 11.68$, $p < .01$, Positive Anticipation, $F(1, 311) = 4.05$, $p < .05$, and Shyness, $F(1, 311) = 5.60$, $p < .05$, and lower on High-intensity Pleasure, $F(1, 311) = 3.92$, $p < .05$. Marginal differences indicated higher scores for girls on Discomfort $F(1, 311) = 3.25$, $p < .10$, and Low-intensity Pleasure, $F(1, 81) = 3.53$, $p < .10$. Significant gender effects in the same directions were obtained from Sample 2 primary caregivers for Fear, $F(1, 81) = 7.22$, $p < .01$, and High-Intensity Pleasure, $F(1, 81) = 10.65$, $p < .01$. In addition, females were rated lower than males on Activity Level, $F(1, 81) = 4.78$, $p < .05$, and marginal differences were observed indicating higher levels of Inhibitory Control, $F(1, 81) = 3.75$, $p < .10$, and Perceptual Sensitivity, $F(1, 81) = 3.46$, $p < .10$ in girls. Secondary caregivers in Sample 2 gave females higher ratings for Cuddliness, Inhibitory Control, Positive Anticipation, and Sociability, Fs(1, 33) = 6.74, 4.62, 7.83, and 5.23, respectively, ps < .05. Marginal effects suggested that girls were higher than boys in Low-intensity Pleasure and Attention Shifting, but lower in Activity Level, Fs(1, 33) = 3.77, 3.20, and 3.06, respectively, ps < .10.

Gender differences were also found for the factor-level scores. Sample 1 caregivers rated girls significantly higher in Negative Affectivity, $F(1, 311) = 5.78$, $p < .05$ and marginally higher in Effortful Control, $F(1, 311) = 3.27$, $p < .10$. Sample 2 primary caregivers perceived females as higher on Negative Affectivity than males, $F(1, 90) = 6.46$, $p < .05$. Secondary caregivers in Sample 2 perceived girls as higher in Effortful Control than boys, $F(1, 30) = 9.56$, $p < .01$.

6. Discussion

The results of this study provide initial support for the reliability and validity of the ECBQ as a measure of finely differentiated and varied aspects of toddler temperament. The 18 scales were internally consistent, and for the large majority of dimensions, raters were both consistent with one another and consistent across time. The factor structure of the instrument was similar to that of fine-grained measures of temperament currently in use with older and younger children. Age and gender differences that emerged from this study were also generally consistent with prior literature.

Although adequate internal consistency was apparent for the large majority of scales at all time points, researchers whose primary interests concern Impulsivity are advised to exercise caution, since this scale performed relatively poorly in both samples, and with both primary and secondary caregivers. It may be the case that speed of response initiation is more context-specific than other dimensions, leading to relatively low intercorrelations among scale items. The prominence of Impulsivity in theory and empirical work regarding temperament (e.g., Buss & Plomin, 1975; Schwebel, 2004) and personality (e.g., Miller, Joseph, & Tudway, 2004), combined with adequate cross-rater and longitudinal stability of this scale, support our decision to retain this scale in the ECBQ.

The factor structure of the ECBQ was consistent across the two samples. In addition, the three factors were highly similar to those that have emerged in analysis of comparable fine-grained measures suitable to infants (e.g., Gartstein & Rothbart, 2003) and older children (Rothbart et al., 2001). This consistency is encouraging when one considers the considerable differences in items comprising scales in the IBQ-R, ECBQ, and CBQ. In addition, the Negative Affectivity, Surgency/Extraversion, and Effortful Control appear comparable to Neuroticism, Extraversion, and Constraint/Conscientiousness constructs identified in empirical investigations of adult personality (e.g., Digman, 1990; Eysenck, 1967; Goldberg, 1990; see Putnam et al., 2001). The inclusion of scales not found on the IBQ-R or CBQ provide additional information regarding the nature of the derived factors. Primary loadings of Sociability on Surgency are consistent with interpretations of this factor as conceptually similar to Extraversion, a trait involving the desire to interact with others (Digman, 1994). The presence of Attention Shifting on the Effortful Control factor supports theory proposed by Posner and Rothbart (1998) concerning an integrated system involved in both sustaining and intentionally reallocating attention. Given the apparent similarity of Motor Activation to Activity Level, one may have expected

Motor Activation to load primarily on Surgency, rather than Negative Affectivity. These findings are, however, consistent with adult personality literature. Zhong and Qian (2005) reported positive correlations between self-reported fidgeting and depression; and Mehrabian and Friedman (1986) found fidgeting to be associated with "unpleasant" temperament.

Also intriguing are scales loading inconsistently across the infant, toddler, and child measures. Putnam et al. (2001) suggested that Perceptual Sensitivity in infancy may be indicative of reactive reward orientation under the influence of a Surgency system, but a marker of children's ability to flexibly engage in quiet activities in older children. In toddlers, however, this scale loads primarily on Negative Affectivity, perhaps reflecting parents' tendencies to recognize toddler awareness of mild stimuli through the child's expressed negativity. Despite considerable similarity in scale items across the toddler and childhood instruments, Positive Anticipation shifts from the Surgency factor on the ECBQ to Negative Affectivity on the CBQ. This may be due to increasing parental expectations for children to control their enthusiasm at older ages. Alternatively, increasing awareness of denied pleasures may lead to closer connections between anticipation of rewards and anger/frustration when these expectations are not met (Putnam et al., 2001).

When summary scores for the factors were created, a negative correlation between Negative Affectivity and Effortful Control was evident. These results are consistent with Ahadi et al. (1993), who found a negative relation between CBQ Negative Affect and Effortful Control for U.S. children, whereas analyses of a Chinese sample indicated Effortful Control was not related to Negative Affect, but rather was inversely associated with Surgency. Ahadi et al. (1993) suggested that willful control may be used in the service of dampening the expression of traits deemed unacceptable in a society. Chen et al. (1998) have demonstrated more favorable attitudes toward children's fear among Chinese parents, in comparison to Canadian. The results of our analyses suggest that other negative emotions may be similarly discouraged among parents from individualist cultures.

Although primary and secondary caregivers often play differing roles (Parke, 1995), interacting with children in different ways and in different contexts, caregiver agreement was substantial for several dimensions, particularly those regarding specific negative affects (e.g., Fear, Frustration). Expressions of negative affect, in addition to behaviors indicative of Activity Level and Inhibitory Control, for which inter-rater reliability was similarly high, may be particularly salient to parents and readily observable. In contrast, because expressions of low-intensity pleasure and indicators of attentional control tend to be more subtle in their overt manifestations, parent ratings of these constructs may be more subjective, and vary as a function of that caregiver's unique experience with the child. The level of, and variability across dimensions in, correspondence between mother and father ratings is comparable to those obtained with the TBAQ (Goldsmith, 1996) and other temperament instruments (see reviews by Rothbart & Mauro, 1990; and Slabach, Morrow, & Wachs, 1991). Multiple hypotheses have been offered to explain the limited convergence of parental ratings of temperament. Mothers and fathers often interact with children at different times and may elicit different behaviors, contributing to discrepancies in the conduct upon which their ratings are based (Bates, 1989; Rothbart & Goldsmith, 1985). Parental personality, response sets such as social desirability and acquiescence, and differential memory for events may also contribute to these inconsistencies (Rothbart & Bates, in press).

Moderate levels of cross-rater agreement, combined with limited convergence of questionnaire and observational data, have led some authors (e.g., Kagan, 1994, 1998) to question the continued use of parent reports. In light of these arguments, it is important to recognize factors contributing to imperfect measurement in laboratory and home observation measures, including those related to characteristics of the rater or experimenter and effects of the measure on child behavior (Rothbart & Bates, 1998; Rothbart & Goldsmith, 1985). In addition, ethical and practical constraints limit the number and type of contexts to which children can be exposed in the laboratory and may not allow for detection of rare but important events. The time and expense involved in generating observational data has resulted in few assessments of test–retest reliability, with these studies suggesting adequate reliability only when multiple assessments are conducted. For instance, Seifer, Sameroff, Barrett, & Krafchuk (1994) achieved satisfactory reliability when aggregating over eight laboratory sessions, but week-to-week reliabilities of their temperament battery ranged from .14 to .36, which they contended were typical for studies of infant behavior. Although both forms of data yield valid information, the limitations of each suggest that researchers utilize both whenever possible (Rothbart & Bates, in press). Relevant procedures from the Lab-TAB (Goldsmith & Rothbart, 1991), in addition to innovative observational protocols to assess traits not assessed in previous studies, will be useful in this regard. Despite the limitations of parent-report instruments, however, it should be noted that recent evidence suggests that parental report of child temperament

may have superior predictive validity relative to other sources of information addressing child temperament, including structured observations (Hart, Field, & Roitfarb, 1999; Pauli-Pott, Mertesacker, Bade, Haverkock, & Beckmann, 2003).

Most of the correlations representing longitudinal stability over 6- and 12-month spans were large in magnitude, and the vast majority of stability coefficients from 18 to 36 months were moderate to large (Cohen, 1988). These values are comparable to those obtained over similar spans in previous studies of temperament (e.g., Earls & Jung, 1987; Guerin & Gottfried, 1994). The Positive Anticipation scale exhibited the lowest stability among both primary and secondary respondents. Tendencies to express excitement over upcoming activities may be more open to socialization influences than other dimensions assessed by the ECBQ. In contrast, Activity Level and aspects of Negative Affectivity were considerably stable, perhaps justifying the universality of these dimensions in multiple theoretical temperament frameworks and measures (Buss & Plomin, 1975; Goldsmith et al., 1987; Rothbart & Mauro, 1990). The inclusion of dimensions assessing Effortful Control is more specific to the regulative processes that are part of our approach, and no other studies to date have documented stability of fine-grained aspects of these processes before the age of three. Although replication utilizing observational data is needed to address the possibility that the consistency revealed in this investigation is not due solely to continuity in parents' views, the current study is the first to demonstrate stability over toddlerhood of parental perceptions of children's abilities to willfully control attention and manage undesirable behavior; and to enjoy calm and affectionate activities.

As expected on the basis of Goldsmith's (1996) report of age-related increases in parent ratings of temperament, our results tentatively suggest increases in multiple traits over toddlerhood, although only Attention Focusing, Discomfort, Inhibitory Control, and Positive Affect increased significantly across all three data sources. At the factor level, Effortful Control exhibited a significant age effect for all samples and both mothers and fathers in Sample 2 perceived increases in Negative Affect. Recent observational findings (e.g., Gerardi-Caulton, 2000; review by Rothbart & Rueda, 2005) converge in support of increasing capability for self-regulation and flexible allocation of attention during the third year of life. It is surprising, however, that children apparently do not effectively use this increased control to reign in their negative affect displays. One potential explanation concerns the shift from externally supported regulation of affect during infancy to increasing expectations of autonomous regulation in toddlerhood. Even if children are becoming more able to control their emotions, these abilities may not be sufficient to counter the decreased support received from parents over this period. The nature of our data suggest an alternative explanation for increases over time in Surgency and Negative Affect, observed in only Sample 2 participants. Parental expectations for, and perception of, increased abilities for the control of behavior may make instances of excessive activity and emotionality more salient to caregivers, leading to higher ratings at later time points.

The apparent presence of normative change, combined with moderate levels of stability, raise important questions regarding the meaning of the scores derived from the ECBQ and whether individual differences may have different meanings at different ages. As noted by Rothbart (1989), aspects of temperament are not static attributes, but emerge gradually through processes involving maturation and experience, and our data suggest both considerable normative change and fairly stable individual differences. Because children differ in maturational rates for most developmental phenomena, it is likely that, to some degree, these scores confound maturational status with individual differences in the underlying systems, potentially leading to underestimates of the longitudinal stability of the actual trait. Alternatively, longitudinal instability over this crucial developmental period may indicate impaired development of self-regulatory abilities, perhaps in response to environmental factors such as parenting.

Consistent with gender differences in maternal perceptions of infants (Gartstein & Rothbart, 2003), primary caregivers in both samples rated female toddlers as higher than males in Fear, but lower in High-intensity Pleasure. Although the early appearance of gender differences in fear can be interpreted as evidence of biological factors, socialization may also be important. Indicating acceptance of fearfulness in girls, Simpson and Stevenson-Hinde (1985) reported better mother–child relationships among socially fearful girls, whereas less fearful boys appeared to be favored. A large body of research (see Zuckerman, 1994) has shown adult males in multiple cultures to be higher than females in sensation seeking, a personality dimension similar to High-intensity Pleasure. Zuckerman (1994) contends that gonadal hormones influence sensation seeking by lowering MAO levels in the brain. The combined results of the current study and Gartstein and Rothbart (2003), who found higher High-intensity Pleasure in male than female infants, suggest that these hormones may have an organizing influence on behavior very early in development.

It is curious that secondary, but not primary, caregivers viewed female toddlers as significantly higher than males in Cuddliness and Sociability. Fathers tend to be more gender-stereotypical than mothers in their activities with children (Turner & Gervai, 1995), and may allow more opportunities for cuddling with daughters than sons (Snow, Jacklin, & Maccoby, 1983). Fathers' greater preference for rough-and-tumble games with boys (Eccles, Freedman-Doan, Frome, Jacobs, & Yoon, 2000) may also limit engagement in the kinds of play likely to elicit demonstrations of sociability from boys. Finally, because fathers spend less time with their daughters than their sons (Manlove & Vernon-Feagans, 2002), making fathers a more scarce resource for daughters than sons, daughters may express greater positive anticipation in activities involving fathers.

7. Summary

The ECBQ was designed to provide a more comprehensive and detailed assessment of temperament than can be obtained through other existing measures appropriate for toddlers (e.g., Fullard, McDevitt, & Carey, 1984; Goldsmith, 1996). This enhanced specificity will likely lead to an increased understanding of developmental processes underlying temperament, in addition to more precise investigations of relations between temperament and other constructs, including behavior problems, parent–child interaction, and school readiness capabilities. The scales comprising the ECBQ are internally consistent, demonstrate satisfactory cross-rater agreement, and are longitudinally stable to an expectable degree. The type of items used in the ECBQ, based on specific responses to specific situations, may also lend advantages over questions that ask parents to generalize over contexts. In addition to potentially reducing rater bias, such items allow for scores that are more clearly interpretable (Goldsmith, 1996) and appropriate for assessing developmental shifts in displays of temperament traits, as evidenced by strong and theoretically anticipated increases in dimensions of effortful control over toddlerhood.

The consistency of our findings between primary and secondary caregivers, and over two separate samples, both cross-sectional and longitudinal, enhances our confidence in the internal validity of our findings. Additional work must be done, however, to establish the concurrent, predictive, and external validity of the instrument. A particularly important step is to ascertain convergence between ECBQ scale scores and standardized laboratory assessments. Another is to investigate longitudinal concordance between the toddler measure and fine-grained measures designed for infants and older children. In addition, although a continent separated the two samples utilized in the current study geographically, the ethnic and socio-economic composition of the samples was highly similar, and additional studies are necessary to determine the usefulness of the measure in more diverse populations.

The degree of specificity afforded by the ECBQ represents an important advance in the measurement of temperament. It is important, however, to realize that this advantage may come with some costs. Due to the large number of traits assessed, relatively few items are used to tap each dimension. In contrast, the longer scales utilized by the TBAQ enhance the internal consistency and increase the domain content of scales from this measure. The TBAQ has also been validated across multiple populations and corresponds to standardized laboratory tasks assessing temperament domains (Goldsmith & Rothbart, 1991). As such, researchers whose main interests concern Activity Level, undifferentiated Pleasure, Social Fearfulness, Anger, or Interest are advised to use the TBAQ, whereas researchers with more differentiated interests, those whose theoretical approach emphasizes regulation in addition to reactivity, and those who desire comparability between their work and empirically derived models of personality (e.g., Goldberg, 1990) will find greater success with the ECBQ.

Acknowledgement

This research was supported by funds allocated to Samuel Putnam and Maria Gartstein from NIMH (U.S.) grant 5 T32 MH1893, awarded to University of Oregon, and from a new faculty award from Bowdoin College, awarded to Samuel Putnam. The authors wish to acknowledge Laura Backen-Jones for her contributions to the interim version of the ECBQ, Elizabeth White for her assistance in data collection, and Hill Goldsmith for his careful work on the TBAQ, which served as both inspiration for, and origin of, the ECBQ. Preliminary findings were presented to the International Society for Infant Studies in April, 2002.

Appendix A. Scale definitions: Early Childhood Behavior Questionnaire

Scale label	Definition
Activity Level	Level (rate and intensity) of gross motor activity, including rate and extent of locomotion.
Attentional Focusing	Sustained duration of orienting on an object of attention; resisting distraction.
Attentional Shifting	The ability to transfer attentional focus from one activity/task to another.
Cuddliness	Child's expression of enjoyment in and molding of the body to being held by a caregiver.
Discomfort	Amount of negative affect related to sensory qualities of stimulation.
Fear	Negative affect related to anticipated pain, distress, sudden events and/or potentially threatening situations.
Frustration	Negative affect related to interruption of ongoing tasks or goal blocking.
High-Intensity Pleasure	Pleasure or enjoyment related to situations involving high intensity, rate, complexity, novelty and incongruity.
Impulsivity	Speed of response initiation.
Inhibitory Control	The capacity to stop, moderate, or refrain from a behavior under instruction.
Low-Intensity Pleasure	Pleasure or enjoyment related to situations involving low intensity, rate, complexity, novelty and incongruity.
Motor Activation	Repetitive small-motor movements; fidgeting.
Perceptual Sensitivity	Detection of slight, low intensity stimuli from the external environment.
Positive Anticipation	Excitement about expected pleasurable activities
Sadness	Tearfulness or lowered mood related to suffering, disappointment, or loss.
Shyness	Slow or inhibited approach and/or discomfort in social situations involving novelty or uncertainty.
Sociability	Seeking and taking pleasure in interactions with others.
Soothability	Rate of recovery from peak distress, excitement, or general arousal.

References

Ahadi, S. A., Rothbart, M. K., & Ye, R. M. (1993). Child temperament in the U.S. and China: Similarities and differences. *European Journal of Personality*, *7*, 359–377.

Almli, C. R., Ball, R. H., & Wheeler, M. E. (2001). Human fetal and neonatal movement patterns: Gender differences and fetal-to-neonatal continuity. *Developmental Psychobiology*, *38*(4), 252–273.

Bates, J. E. (1989). Applications of temperament concepts. In G. A. Kohnstamm, J. E. Bates, & M. K. Rothbart (Eds.), *Temperament in childhood* (pp. 322–355). New York: Wiley.

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

Campbell, D. W., & Eaton, W. O. (1999). Sex differences in the activity level of infants. *Infant and Child Development*, *8*, 1–17.

Carey, W. B., & McDevitt, S. C. (1978). Revision of the infant temperament questionnaire. *Pediatrics*, *61*, 735–739.

Chen, X., Hastings, P. D., Rubin, K. H., Chen, H., Cen, G., & Stewart, S. L. (1998). Child-rearing attitudes and behavioral inhibition in Chinese and Canadian toddlers: A cross-cultural study. *Developmental Psychology*, *34*, 677–686.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cote, S., Tremblay, R. E., Nagin, D., Zoccolillo, M., & Vitaro, F. (2002). The development of impulsivity, fearfulness, and helpfulness during childhood: Patterns of consistency and change in the trajectories of boys and girls. *Journal of Child Psychology and Psychiatry*, *43*(5), 609–618.

DeVellis, R. R. (1991). *Scale development: Theory and applications*. Thousand Oaks, CA: Sage Publications, Inc.

Digman, J. D. (1994). Child personality and temperament: Does the five-factor model embrace both domains? In C. F. Halverson & G. A. Kohnstamm (Eds.), *Developing structure of temperament and personality from infancy to adulthood* (pp. 323–338). Hillsdale, NJ: Lawrence Erlbaum Associates.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, *41*, 417–440.

Earls, F., & Jung, K. G. (1987). Temperament and home environment characteristics as causal factors in the early development of childhood psychopathology. *Journal of the American Academy of Child and Adolescent Psychiatry*, *26*, 491–498.

Eaton, W. O., & Enns, L. R. (1986). Sex differences in human motor activity level. *Psychological Bulletin*, *100*, 19–28.

Eccles, J. S., Freedman-Doan, C., Frome, P., Jacobs, J., & Yoon, K. S. (2000). Gender-role socialization in the family: A longitudinal approach. In T. Eckes & H. M. Trautner (Eds.), *Developmental social psychology of gender* (pp. 333–360). Mahwah, NJ: Lawrence Erlbaum Associates.

Eiden, R. D., Edwards, E. P., & Leonard, K. E. (2004). Predictors of effortful control among children of alcoholic and nonalcoholic fathers. *Journal of Studies on Alcohol*, *65*, 309–319.

Eysenck, H. J. (1967). *The biological basis of personality*. Springfield, IL: Thomas.

Fullard, W., McDevitt, S. C., & Carey, W. B. (1984). Assessing temperament in one- to three-year-old children. *Journal of Pediatric Psychology*, *9*, 205–217.

Gartstein, M. A., & Rothbart, M. K. (2003). Studying infant temperament via the Revised Infant Behavior Questionnaire. *Infant Behavior and Development*, *166*, 1–23.

Gerardi-Caulton, G. (2000). Sensitivity to spatial conflict and the development of self-regulation in children 24–36 months of age. *Developmental Science*, *3*, 397–404.

Goldberg, L. (1990). An alternative "description of personality": The big-five factor structure. *Journal of Personality and Social Psychology*, *59*, 1216–1229.

Goldsmith, H. H., Buss, A. H., Plomin, R., Rothbart, M. K., Thomas, A., Chess, S., et al. (1987). Roundtable: What is temperament? *Child Development*, *58*, 505–529.

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 J. Strelau & A. Angleitner (Eds.), *Explorations in temperament: International perspectives on theory and measurement* (pp. 249–272). New York: Plenum Press.

Guerin, D. W., & Gottfried, A. W. (1994). Temperamental consequences of infant difficultness. *Infant Behavior and Development*, *17*, 413–421.

Hart, S., Field, T., & Roitfarb, M. (1999). Depressed mothers' assessments of their neonates' behaviors. *Infant Mental Health Journal*, *20*(2), 200–210.

Kagan, J. (1994). *Galen's prophecy: Temperament in human nature*. New York: Basic Books.

Kagan, J. (1998). Biology and the child. In W. S. E. Damon & N. V. E. Eisenberg (Eds.), *Handbook of child psychology: Vol 3 Social, emotional and personality development* (5th edition; pp. 177–235). New York: Wiley.

Kochanska, G., DeVet, K., Goldman, M., Murray, K., & Putnam, S. P. (1994). Maternal reports of conscience development and temperament in young children. *Child Development*, *65*, 852–868.

Kochanska, G., & Knaack, A. (2003). Effortful control as a personality characteristic of young children: Antecedents, correlates, and consequences. *Journal of Personality*, *71*, 1087–1112.

Kusanagi, E., Chen, S., & Hoshi, N. (2000). Psychometric properties of the Toddler Behavior Assessment Questionnaire applied to Japanese children. *Psychologia*, *43*, 262–274.

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.

Manlove, E. E., & Vernon-Feagans, L. (2002). Caring for infant daughters and sons in dual-earner households: Maternal reports of father involvement in weekday time and tasks. *Infant and Child Development*, *11*, 305–320.

Martin, R. P., Wisenbaker, J., Baker, J., & Huttunen, M. O. (1997). Gender differences in temperament at six months and five years. *Infant Behavior and Development*, *20*, 339–347.

Maziade, M., Boudreault, M., Thivierge, J., Caperaa, P., & Cote, R. (1984). Infant temperament: SES and gender differences and reliability of measurement in a large Quebec sample. *Merrill-Palmer Quarterly*, *30*, 213–226.

Mehrabian, A., & Friedman, S. L. (1986). An analysis of fidgeting and associated individual differences. *Journal of Personality*, *54*, 406–429.

Miller, E., Joseph, S., & Tudway, J. (2004). Assessing the component structure of four self-report measures of impulsivity. *Personality and Individual Differences*, *37*, 349–358.

Norusis, M. J. (1994). *SPSS professional statistics 6.1*. Chicago: SPSS, Inc.

Parke, R. D. (1995). Fathers and families. In M. H. Bornstein (Ed.), *Status and social conditions of parenting: volume 3*. *Handbook of parenting* (pp. 27–63). Hillsdale, NJ: Lawrence Erlbaum Associates.

Pauli-Pott, U., Mertesacker, B., Bade, U., haverkock, A., & Beckmann, D. (2003). Parental perceptions and infant temperament. *Infant Behavior and Development*, *26*(1), 27–48.

Posner, M. I., & Rothbart, M. K. (1998). Summary and commentary: Developing attentional skills. In J. E. Richards (Ed.), *Cognitive neuroscience of attention: A developmental perspective* (pp. 317–323). Mahwah, NJ: Erlbaum.

Posner, M. I., & Rothbart, M. K. (2000). Developing mechanisms of self-regulation. *Development and Psychopathology*, *12*, 427–441.

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.

Reznick, J. S., Gibbons, J. L., Johnson, M. O., & McDonough, P. M. (1989). Behavioral inhibition in a normative sample. In J. S. Reznick (Ed.), *Perspectives on behavioral inhibition* (pp. 25–49). Chicago: University of Chicago Press.

Rothbart, M. K. (1981). Measurement of temperament in infancy. *Child Development*, *52*, 569–578.

Rothbart, M. K. (1989). Temperament and development. In G. A. Kohnstamm, J. E. Bates, & M. K. Rothbart (Eds.), *Temperament in childhood* (pp. 187–248). New York: Wiley.

Rothbart, M. K. (1994). Emotional development: Changes in reactivity and self-regulation. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 369–372). New York: Oxford University Press.

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. (1998). Temperament. In W. Damon (Series Ed.) & N. Heisenberg (Vol. Ed.), *Handbook of child psychology, vol. 3, Social, emotional, and personality development* (5th ed.; pp. 105–176). New York: Wiley.

Rothbart, M. K., & Bates, J. E. (in press). Temperament in children's development. In W. Damon, R. Lerner, & N. Eisenberg (Eds.), *Handbook of child psychology volume 3, Social, emotional, and personality development* (6th ed.). 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: volume 1* (pp. 37–86). New York: Guilford.

Rothbart, M. K., & Goldsmith, H. H. (1985). Three approaches to the study of infant temperament. *Developmental Review*, *5*, 237–260.

Rothbart, M. K., & Mauro, J. A. (1990). Questionnaire approaches to the study of infant temperament. In J. W. Fagen & J. Colombo (Eds.), *Individual differences in infancy: Reliability, stability, and prediction* (pp. 411–429). Hillsdale, NJ: Erlbaum.

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.

Ruff, H. A., & Lawson, K. R. (1990). Development of sustained, focused attention in young children during free play. *Developmental Psychology*, *26*, 85–93.

Salinas, C. G., Montesinos, M. D., & Carnicero, J. A. C. (1999). Adaptación a población española del "Toddler Behavior Assessment Questionnaire" para la medida del temperamento en la infancia. *Metodologia de las Ciencias del Comportamiento*, *1*, 207–218.

Seifer, R., Sameroff, A. J., Barrett, L. C., & Krafchuk, E. (1994). Infant temperament measured by multiple observations and mother report. *Child Development*, *65*, 1478–1490.

Schwebel, D. C. (2004). Temperamental risk factors for children's unintentional injury: the role of impulsivity and inhibitory control. *Personality and Individual Differences*, *37*, 567–578.

Simpson, A. E., & Stevenson-Hinde, J. (1985). Temperamental characteristics of three- to four-year-old boys and girls and child-family interactions. *Journal of Child Psychology and Psychiatry*, *26*, 43–53.

Slabach, E. H., Morrow, J., & Wachs, T. D. (1991). Questionnaire measurement of infant and child temperament: Current status and future directions. In J. Strelau & A. Angleitner (Eds.), *Explorations in Temperament: International Perspectives on Theory and Measurement* (pp. 205–234). New York: Plenum.

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, *12*, 102–111.

Snow, M. E., Jacklin, C. N., & Maccoby, E. E. (1983). Sex-of-child differences in father-child interaction at one year of age. *Child Development*, *54*, 227–232.

Stevens, G., & Featherman, D. L. (1981). A revised socioeconomic index of occupational status. *Social Science Research*, *10*, 573–585.

Turner, P. G., & Gervai, J. (1995). A multidimensional study of gender typing in preschool children and their parents: Personality, attitudes, preferences, behavior, and cultural differences. *Developmental Psychology*, *31*, 759–772.

Van Bakel, H. J. A., & Riksen-Walraven, J. M. (2004). AQS security scores: What do they represent? A study in construct validation. *Infant Mental Health Journal*, *25*, 175–193.

Vaughn, B. E., Kopp, C. B., & Krakow, J. B. (1984). The emergence and consolidation of self-control from eighteen to thirty months of age: Normative trends and individual differences. *Child Development*, *55*, 990–1004.

Zhong, J., & Qian, M. (2005). A study of development and validation of Chinese Mood Adjective Check List. *Chinese Journal of Clinical Psychology*, *13*, 8–13.

Zuckerman, M. (1994). *Behavioral expressions and biosocial bases of sensation seeking*. New York: Cambridge University Press.