Journal of Research in Personality 52 (2014) 6–12

Contents lists available at ScienceDirect

Journal of Research in Personality

journal homepage: www.elsevier.com/locate/jrp

Structure and convergent validity of children's temperament traits as assessed by experimenter ratings of child behavior

Lisa N. Vroman *, Sharon L. Lo, C. Emily Durbin

Department of Psychology, Michigan State University, United States

ARTICLE INFO

Article history:
Available online 13 June 2014

Keywords:
Temperament
Structure
Experimenter ratings

ABSTRACT

The structure of child temperament traits has been explored primarily using informant report. Less is known about temperament structure assessed by alternative methods, such as laboratory assessments. We report on the structure of child traits assessed by experimenter ratings of child behavior during laboratory tasks, and their convergent and discriminant validity with objectively coded and parent reported child traits. The results indicate a three-factor solution (Positive Emotionality, Negative Emotionality, and Effortful Control) fit the data best, with convergent and discriminant validity between experimenter ratings and objective coding of child behavior and parent report. The results suggest that experimenter ratings conducted after a laboratory visit provides an efficient and economical alternative or adjunct to conducting objective coding of the laboratory tasks.

1. Introduction

Individual differences in emotional reactivity and self-regulation, or temperament, have long been identified as among the earliest emerging biobehavioral differences in children (Rothbart & Derryberry, 1981). Several theoretical traditions have emerged in developmental research to describe the main dimensions underlying these early differences in temperament (i.e., Buss & Plomin, 1984; Rothbart, 1981; Thomas & Chess, 1977). Most contemporary temperament models propose a multidimensional structure for temperament traits in early childhood through adolescence, with models converging to suggest that the primary dimensions concern individual differences in the experiential, expressive, and motivational components of positive and negative emotions, and in dimensions of behavioral or Effortful Control (e.g., De Pauw & Mervielde, 2010; Goldsmith et al., 1987; Halverson et al., 2003; Rothbart, Ahadi, Hershey, & Fisher, 2001).

Most of the empirical evidence regarding the structure of child temperament has relied on parent questionnaire methods, with a small number of studies using teacher reports (e.g., De Pauw, Mervielde, & Van Leeuwen, 2009; Digman & Shmelyov, 1996; Presley & Martin, 1994). These examinations of informant reports of child temperament have reliably uncovered at least three superfactors: Positive Emotionality/Surgency (PE), Negative Emotionality (NE), and Effortful Control (EC) (Ahadi, Rothbart, & Ye, 1993; Casalin, Luyten, Vliegen, & Meurs, 2012; Rothbart, 2007; Rothbart et al., 2001). PE is generally described as reflecting positive mood, engagement with the environment, and sociability. NE generally refers to individual differences in the frequency and intensity of experiencing negative emotions, including anger/frustration, sadness, and fear. EC is generally described as reflecting aspects of behavioral control, including control of cognitive resources as well as of impulses or behavioral tendencies. For example, Rothbart et al. (2001) assessed the structure of temperament in children 3–7 years of age assessed via parent-report on the Children's Behavior Questionnaire (CBQ). Factor analyses revealed three temperament superfactors: PE, NE, and EC. These results are consistent with an earlier investigation completed by Ahadi et al. (1993) in which the structure of temperament was also investigated via parent-report on the Children's Behavior Questionnaire in both a U.S. and Chinese sample, and a similar three-factor solution was obtained.

An advantage of these three-factor models of temperament is that they are theoretically consistent with Tellegen's three-factor model of personality in adults, consisting of PE, NE, and Constraint (Tellegan, 1985), wherein Constraint is similar to EC. However, other studies using parent or teacher questionnaires have reported that anywhere from three to six factors fit the data best, with additional traits generally converging around the subdomains of sociability, activity/impulsivity, and the division of anger and fear into separate factors (rather than both collapsed into one broad NE factor) (e.g. De Pauw et al., 2009; Halverson et al., 2003; Presley &

* Corresponding author. Address: Department of Psychology, Michigan State University, East Lansing, MI 48824, United States.
E-mail address: vromanli@msu.edu (L.N. Vroman).

http://dx.doi.org/10.1016/j.jrp.2014.06.002
0092-6566/

Martin, 1994). Thus, it is important to utilize multiple methods of assessing temperament to expand our understanding of how early-appearing dispositions cohere into higher-order structures. Comparing results across distinct methods will test the validity of structural results, and may also inform attempts to understand the development of individual differences and biological processes contributing to their development (Shiner et al., 2012). As each method of assessing child temperament has strengths and limitations, the use of multiple methods will help clarify the nature of child temperament. We propose that explorations of temperament structure using methods other than parent report may provide particularly useful incremental knowledge.

Parent-report measures provide ecologically valid information regarding parent perceptions of child behavior; however, parent-report is not without limitations. As noted by several researchers (e.g., Kagan, Snidman, McManis, Woodward, & Hardway, 2002), there are several factors that may weaken the validity of parent-report of child temperament. First, parents who have not had much experience with children are at a disadvantage when responding to questionnaire items that ask them to judge their child's behavior (and hence traits) relative to norms. Second, parent-report of child temperament is likely comprised of both objective and subjective influences (Stifter, Willoughby, & Towe-Goodman, 2008). For example, parents' perceptions of their child's temperament may be influenced by his/her own emotional state. Third, there is evidence that parent-reports of child behavior are biased by their own psychopathology and personality (Durbin & Wilson, 2012).

Laboratory measures of child temperament represent an important complement to informant-report measures for assessing individual differences in child traits. First, they are less influenced by subjective biases evident in parent-report measures, wherein parent characteristics are difficult to disentangle from their perceptions of child traits. Second, they provide access to a fine-grained sample of behaviors that can be mined for evidence of multiple traits, as coding from videotaped tasks can be designed to measure a multitude of traits without the need for recall of specific child behaviors. Third, because they use standardized probes, differences across children in their responses are more easily observed and responses to important, but infrequently encountered stimuli, such as those that may elicit fear, can be more readily assessed.

There is a growing literature employing laboratory tasks to assess individual differences in child traits (e.g., Carlson & Wang, 2007; Dennis, Brotman Miller, Huang, & Kiely Gouley, 2007; Durbin, 2010; Kochanska, Murray, Jacques, Koenig, & Vandegeest, 1996; Willoughby, Wirth, & Blair, 2012). Evidence regarding the structure of traits assessed in this manner is slim, but suggests that coding of child behaviors produces evidence for three broad temperament dimensions (PE, NE, and EC) that are similar to those described in the literature on parent-report. In a recent investigation, Dyson, Olino, Durbin, Hill Goldsmith, and Klein (2011) reported on the factor structure of temperament in preschoolers, whose traits were assessed by a battery of lab tasks, including several drawn from the Laboratory Temperament Battery (Lab-TAB; Goldsmith, Reilly, Lemery, Longley, & Prescott, 1995). The authors uncovered a five-factor model: Sociability, Positive Affect/Interest, Dysphoria, Fear/Inhibition, and Constraint. This five-factor solution is similar, but distinct from, other models that have examined the structure of child temperament via parent-report. For example, the results overlap with Rothbart et al.'s (2001) three-factor model of temperament (i.e., PE, NE, and EC) derived from parent-report on the CBQ, but with a few subtle distinctions. Data from laboratory tasks split the broad PE factor into sociability and positive affect, and the NE factor into fear and sadness/anger. In a similar investigation, Kotelnikova, Olino, Mackrell, Jordan, and Hayden (2013) assessed the structure of temperament in middle childhood by administering a battery of seven laboratory tasks to a community sample of 205 seven-year-old children. The data supported a four-factor model comprised of Positive Emotionality/sociability, disinhibition/anger, fear/behavioral inhibition, and sadness. Here, the authors identified a factor akin to PE as defined in 3-factor models. However, NE was split into fear/behavioral inhibition, and sadness. An EC factor was not extracted in this sample, which may be attributable to the fact that lab tasks designed to elicit this trait were not included in the battery.

Given that most evidence suggests the level of convergence between different methods of assessing child temperament is low (e.g., Durbin, Hayden, Klein, & Olino, 2007; Majdandzic & van den Boom, 2007), it is important to evaluate not only the similarity of trait structure across methods, but also their areas of convergence and divergence. The low-to-moderate convergence across these multiple methods of assessment suggests that the use of different measurement approaches may contribute to the discrepancies among recovered trait structures. In an examination of infant temperament structure assessed via both parent-report on the Infant Behavior Questionnaire (IBQ; Rothbart, 1981) and laboratory measures, modest-to-moderate convergence between the two methods was observed across the PE, NE, and EC temperament superfactors (Rothbart, Derryberry, & Hershey, 2000). In toddlerhood and preschool-aged children, the convergence between behavioral and caregiver measures of temperament (on the CBQ) is also moderate (Kochanska et al., 1996). Several potential sources of difference between laboratory methods employing objective coding of child behavior and parent-report questionnaires likely contribute to lower convergence. First, coders typically see only a small sample of child behavior in a single (or a few) contexts, whereas parents see behaviors across a variety of contexts, particularly recurring contexts, allowing for inferences about their child's behavioral style across different situations and in response to similar and dissimilar stimuli. Second, coders have the advantage of videotaped samples that reduce their memory burden and allow for minute examination of behaviors that are not possible during live interaction. Third, coders do not have any relationship with the children they code, such that their stance towards a child's behavior is more neutral than for someone who engages in ongoing interactions with that child (i.e., their parent), and whose relationship to the child has both a history and a deep personal meaning.

In the current project, we examined the factor structure and convergent validity of ratings of child traits made by experimenters who conducted laboratory assessments with children. This approach offers an interesting comparison to both coding and parent-report methods. These ratings are similar to traditional coding of laboratory tasks in that they are made based on the behaviors exhibited by a child during structured lab tasks and are completed by people with no prior experience or long-standing relationship with the child. However, they are similar to parent-report measures in that they require an aggregation of perceptions recalled across a longer sample of behavior (a 2-hour visit), and the reporter (the experimenter) has an interactive role with the child as they engage him or her in the laboratory tasks. Finally, if their structural and convergent validity were supported, they also could potentially serve an economical, substitute for or supplement to data from objective coding measures, as experimenter ratings can be collected immediately after a laboratory assessment and are readily scored.

We examined the factor structure, convergent, and discriminant validity of experimenter ratings of child temperament traits collected following a battery of emotion-eliciting laboratory tasks in a sample of 168 young children. Experimenter ratings were compared to parent-reports of child traits and scores derived from objective coding of child behavior. Based on previous research using parent-report and laboratory methods, it was predicted that

at least three factors would emerge, comprising the superfactors of PE, NE, and EC. We also predicted that experimenter ratings of child temperament traits would demonstrate convergent and discriminant validity with objective coding of child behavior and weaker associations with parent-report of child traits.

2. Methods

2.1. Participants

Child participants (N = 168) were recruited from the greater Chicago, Illinois area for a study of child temperament.[1] Children were between the ages of 3 and 7 years. The mean age was 55.7 months (SD = 12.1), and 45.8% were girls. Mothers were between the ages of 23 and 52 years (M = 36.8, SD = 5.1), and fathers were between the ages of 23 and 57 years (M = 38.3, SD = 6.9). Data and race, ethnicity, and family income were provided by 72.1% of mothers and by 70.2% of fathers. Of those, the ethnic composition was as follows: Caucasian/White (77.4%), Hispanic/Latino (10.1%), African American/Black (8.0%), Asian (5.9%), other (3.1%), and bi- or multiracial (2.8%); categories do not sum to 100% because participants could endorse multiple categories. Yearly family income ranged from $21,000 to greater than $100,000; 18.4% reported income less than $41,000. Children were administered the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 1997) to assess their level of receptive language skills (M = 106.62, SD = 15.07).

2.2. Laboratory assessment of child temperament

Children completed a battery of 10 emotion-eliciting laboratory tasks designed to assess temperamental differences positive and Negative Emotionality (PE and NE) and Effortful Control (EC). Six tasks were from the Preschool version of the Laboratory Temperament Assessment Battery (Lab-TAB; Goldsmith et al., 1995), and four tasks were either modified from the Lab-TAB or newly developed for work with this sample. A female experimenter administered all laboratory tasks and tasks were administered in the same order across participants. The parent(s) of each child was present for all but three tasks (noted below). For each of the tasks the parent was present for, he/she was asked to remain neutral. Breaks of approximately 2–4 min were taken in between each episode to allow children to return to a baseline emotional state prior to beginning the next task. The 10 tasks broadly assessed PE, NE, and EC.

2.2.1. Exploring new objects

The child was left to explore the room, which contained novel and ambiguous stimuli. The stimuli included a tunnel connected to a tent, an animal crate containing toy mice, a remote-controlled spider, a wooden box containing sticky "worms", and a plastic skull hidden under a red cloth. The experimenter returned after 4.5 min and asked the child to touch each object.

2.2.2. Making a t-shirt

The child decorated a t-shirt with puff paints and fabric markers. The child was allowed to take the t-shirt home as a gift at the end of the lab visit.

2.2.3. Disappointing toy

The child was shown two pictures of appealing toys (puppets and remote-controlled cars) and one of an unappealing toy (plastic watering can) and asked to select his/her favorite. The experimenter then returned with the unappealing toy instead of the child's choice and left the child to play with the toy for 2 min. The experimenter then returned with the child's favorite toy and the child and experimenter played with the toy together.

2.2.4. Stranger approach

The child was left alone briefly in the testing room. A male research assistant entered the room and spoke to the child in a neutral voice while gradually approaching the child and engaging in a scripted conversation. The child's parent(s) were not in the room for this task. This task was not conducted for nine children because of lack of availability of a male research assistant.

2.2.5. Dress up

The child and experimenter played with dress-up costumes (e.g., fireman's jacket and hat; feather boa and necklace; and doctor's jacket and pretend stethoscope). The experimenter took a Polaroid picture of the child in his/her costume and they watched the picture develop together.

2.2.6. Transparent box

The experimenter locked an appealing toy in a clear plastic box and left the child with an incorrect set of keys to open the lock and play with the toy. After 3 min, the experimenter returned with the correct set of keys and explained that she accidentally gave the child the wrong set of keys. The child was then allowed to open the box and play with the toy.

2.2.7. Popping bubbles

The experimenter made bubbles with a bubble-shooting toy and encouraged the child to pop the bubbles.

2.2.8. Impossibly perfect green circles

The experimenter repeatedly asked the child to draw green circles on a piece of paper while mildly criticizing each circle. After 2 min, the experimenter positively commented on the child's circles.

2.2.9. Pop-up snakes

The experimenter showed the child what looked to be a can of potato chips, but instead contained coiled spring snakes. The experimenter demonstrated the trick and encouraged the child play the trick on his/her parent. The parent was not present for the first half of the task.

2.2.10. Box empty

The experimenter gave the child a gift-wrapped empty box under the pretense that an appealing gift was inside. The child was left alone for 2.5 min to discover that the box is empty. The experimenter then returned with two small toys for the child to take home, explaining that she accidentally forgot to put the toys inside the box. The parent was not present for this task.

2.3. Experimenter rating of children's temperament

Following the laboratory visit, the primary experimenter used the Child Behavior Scale (Gagne, Van Hulle, Aksan, Essex, & Goldsmith, 2011) to rate the child's behavior during the visit on 24 different items: overall positive affect, overall negative affect, energy, adaptation to change in test materials, interest in test materials and stimuli, initiative with tasks, exploration of objects, attention to tasks, persistence in attempting to complete tasks, enthusiasm towards tasks, fear, frustration with inability to complete tasks, social engagement with child tester, social engagement with parent, cooperation with child tester, cooperation with

[1] Other papers using this data set have reported on correlates of lab-assessed child temperament traits: Durbin (2010), Durbin and Wilson (2012), Olino, Durbin, Klein, Hayden, and Dyson (2013), and Wilson and Durbin (2012).

parent, hyperactivity, shyness, prone to anger/irritability, prone to sadness, contentment, exuberance, anticipatory positive affect, and impulsivity. Each item was rated on a five-point Likert scale: 1 = 1 behavior rarely or never exhibited; 2 = slight or ambiguous signs of the behavior; 3 = unambiguous tendency toward behavior; 4 = behavior exhibited to a typical degree; 5 = behavior exhibited to a high degree. In making these ratings (see Table 1), the experimenter considered not only the child's behavior during each structured task, but also all other behavior observed during the course of the visit (i.e., upon arrival to the lab, during free play that occurred in between tasks, and prior to leaving the lab). The experimenter made each rating immediately following the laboratory assessment.

2.4. Objective rating of children's temperament traits

Each episode was coded using a global coding system validated in earlier studies examining child temperament (Durbin, 2010; Durbin, Klein, Hayden, Buckley, & Moerk, 2005; Durbin et al., 2007). Coders were trained graduate and undergraduate students. Episodes were coded by one rater, with the exception of a subset of videos coded by multiple raters to index reliability. To assess emotionality traits (PE and NE), coders recorded all instances of discrete emotional states (i.e., happiness, sadness, fear, anger, surprise), as demonstrated by facial, vocal, and bodily indicators of each of the aforementioned emotions. The AFFEX coding system was used to code the intensity of facial expressions (Izard, Dougherty, & Hembree, 1983) according to three levels: (1) ambiguous or low intensity (expressions of low intensity in one facial region (i.e., eyes or mouth)); (2) moderate intensity (expression definitely present in at least one facial region); and (3) high intensity (expression definitely present in both facial regions). Intensity of vocal and bodily expressions (low, moderate, high) was indicated on a three-point scale. For vocalizations, intensity level was determined by the extent to which tone and/or content conveyed the emotion. For bodily expressions, intensity was defined by the magnitude and/or salience of the bodily movements. A weighted average of these scores was created by converting intensity level to the following scale: 1 = low intensity, 2 = moderate intensity, 3 = high intensity, to yield weighted composite scores for happiness, anger, fear, and sadness. Reliability of coding was indexed on a subsample of participants ($N = 15$). Scores were aggregated across all 10 laboratory tasks. Intraclass correlation coefficients (two-way random, absolute agreement; Shrout & Fleiss, 1979) for total expression counts (aggregated across all 10 tasks) were as follows: PE (.90), fear (.66), sadness (.79), and anger (.81).

Coders also completed global ratings of the child's behavior to index the child's level of the following traits: interest/engagement, activity level, anticipatory positive affect, passivity versus initiative, sociability, compliance, and behavioral control versus impulsivity. Each of these behaviors was rated on a four-point Likert scale: 0 = low, 1 = moderate, 2 = moderate-to-high, and 3 = very high, based on all behaviors observed during the full length of the lab task. Interest/engagement ratings were based on the child's degree of persistence, engrossment, and enjoyment in the task. Activity level ratings were based on the child's degree of movement, including speed and vigor in manipulation of objects and movement around the room. Anticipatory PE was based on the child's positive behavioral and emotional response in anticipation of a positive event or reward. Initiative ratings were based on the child's degree of passivity or assertiveness in interactions with the experimenter and parent. Sociability ratings were based on the child's interest in and pursuit of social interaction with the experimenter or parent. Compliance ratings were based on the severity of the child's deliberate unwillingness to comply with the experimenter's or parent's suggestions or commands. Impulsivity ratings were based on the child's tendency towards impatience and impulsivity, in contrast to planful and deliberate behavior. Interclass correlation coefficients for aggregates of these traits (across all 10 tasks) ranged from .65 (engagement) to .94 (activity).

This coding system was used in a previous investigation with this sample, which revealed a three-factor structure of child temperament as assessed by objective coding of the battery of lab tasks (Wilson & Durbin, 2012). The authors conducted a principal-components analysis with a Varimax rotation of the 10 coded temperament traits, revealing three higher order temperament dimensions: PE (loadings for positive emotions, positive anticipation, sociability, engagement), NE (fear, sadness, anger), and EC (compliance, low activity level, low impulsivity).

Table 1
Exploratory factor analysis with child behavior scale variables.

Trait	PE	EC	NE
Overall positive affect	**.81**	.19	−.16
Energy	**.80**	−.06	−.25
Exuberance	**.77**	.05	−.25
Social engagement with child tester	**.76**	.22	−.17
Shyness	**−.75**	−.10	.01
Enthusiasm toward tasks	**.75**	.40[a]	−.14
Initiative with tasks	**.70**	.06	−.38
Anticipatory positive affect	**.62**	−.07	−.02
Exploration of objects	**.62**	.16	−.54[a]
Interest in test materials and stimuli	**.62**	.51[a]	−.17
Social engagement with parent	**.42**	−.14	.21
Cooperation with child tester	.09	**.88**	−.29
Attention to tasks	.12	**.83**	−.27
Cooperation with parent	.17	**.74**	−.07
Adaptation to change in test materials	.10	**.74**	−.37
Impulsivity	.36	**−.68**	−.02
Contentment	.53[a]	**.66**	−.36
Hyperactivity	.47[a]	**−.66**	−.07
Persistence in attempting to complete tasks	.29	**.62**	−.27
Prone to anger/irritability	.02	**−.60**	.47[a]
Frustration with inability to complete tasks	.07	**−.48**	.38
Prone to sadness	−.17	−.23	**.81**
Fear (refers to reactions to objects/situations, not shyness)	−.20	−.10	**.77**
Overall negative affect	−.25	−.53[a]	**.68**

Note. Items loading onto each factor are in bold.
[a] Cross-Loadings ⩾ .40.

2.5. Parent-report of child temperament

Participating mothers and fathers completed the Children's Behavior Questionnaire (CBQ; Rothbart et al., 2001), which was designed to measure temperament in children aged 3–7 years. Parents were given the CBQ and a pre-stamped envelope at the conclusion of the lab visit and asked to complete the questionnaire at home and return by mail. The CBQ consists of 195 items rated on a 1 (extremely untrue) to 7 (extremely true) Likert scale and includes subscales tapping the higher-order dimensions of Surgency/Positive Emotionality, Negative Emotionality, and Effortful Control. CBQ scores were computed separately for maternal and paternal report. CBQ scales have shown adequate internal consistency and good test-retest reliability (Rothbart et al., 2001). In this sample, Cronbach's alphas for the three scales ranged from .73 (maternal report on Surgency) to .89 (paternal report on EC). The alphas for each scale (reported separately for maternal and paternal report) are reported in Table 2. Eighty percent of mothers and 79% of fathers completed the CBQ. Given the high degree of intercorrelation between maternal and paternal report on the three subscales (PE = .73, NE = .59, EC = .50), we aggregated maternal and paternal reports, except for cases in which only one parent provided data ($n = 8$). No significant differences were observed on

any of the experimenter ratings of child behavior between the children for whom we did and did not receive maternal ratings on the CBQ.

3. Results

We report on the structure of experimenter ratings of child behavior, as well as bivariate associations between factor scores derived from experimenter ratings and two other measures, objective coding of child traits and parent-report.

3.1. Exploratory factor analyses

Principal axis factor analysis (PAF) with an Oblimin (oblique) rotation was conducted on the experimenter ratings of child behavior. We selected an oblique rotation, as we wished to allow the emergent factors to correlate, as common dimensions of temperamental differences in children are unlikely to be orthogonal to one another. The PAF factor criterion was evaluated against the following criteria: (a) eigenvalue > 1.00 rule (Kaiser–Guttman criterion), (b) scree test (Gorsuch, 1983), (c) the configuration accounted for a minimum of 50% of the total variance (Streiner, 1994), and (d) a minimum of three variables per factor were required to identify common factors (Anderson & Rubin, 1956; Comrey, 1988). Variables were considered meaningful when their factor loadings were greater than .40. Based on these criteria and classic parallel analysis, a three-factor structure fit the data best, explaining 58.9% of the variance. When the data were constrained to fit a two-factor model, 51.6% of the variance was explained. When the data were constrained to fit a four-factor model, 63.2% of the variance was explained. However, the factor loadings were not as interpretable as the three-factor solution, as the NE factor appeared to be divided into two factors related to fearful/sad and angry/irritable affect, respectively. We determined that these two factors were best collapsed into one general NE factor. We also conducted a principal components analysis with a Varimax rotation to assess whether similar results were obtained with an orthogonal rotation. Here, a three-factor structure also fit the data best, explaining 58.9% of the variance. Convergence across both methods was high, with congruence coefficients ranging from .99 to 1.00. Thus, a three-factor solution was determined to fit the data best.

The three factors were named according to the items that loaded onto each factor. Factor 1 was defined by loadings of overall positive affect, energy, interest in test materials/stimuli, initiative with tasks, exploration of objects, enthusiasm toward tasks, social engagement with child tester and parent, exuberance, anticipatory positive affect, and a negative loading on shyness. Factor 1 appeared to reflect a broad PE dimension encompassing both its mood, motivation, and sociability elements. Factor 2 was defined by loadings of adaptation to change in test materials, attention to tasks, persistence in attempting to complete tasks, cooperation with child tester and parent, and contentment, and negative loadings for frustration with inability to complete tasks, hyperactivity, anger/irritability, and impulsivity. Factor 2 related best to EC, including elements of adaptability, cognitive control, and low levels of behavior problems related to low EC; it also included anger and frustration, traits that sometimes load on NE factors (Rothbart et al., 2001; Wilson & Durbin, 2012). Finally, the third factor was defined by loadings of overall negative affect, fear, and sadness, and therefore appeared to reflect NE.

3.1.1. Cross-loadings

Six cross loadings were noted. The two items of overall negative affect and proneness to anger/irritability both loaded positively onto the NE factor (.68 and .47, respectively) and negatively onto the EC factor (−.53 and −.60, respectively), indicating that these dimensions correlated both with other negative emotions and with elements of behavioral control. The contentment and interest in test materials and stimuli items both loaded positively onto the factors PE (.53 and .62, respectively) and EC (.66 and .51, respectively); contentment (or low intensity PA) has been found to load onto an EC factor in a previous investigation of parent reports on the CBQ (Rothbart et al., 2001). Hyperactivity loaded positively on the PE factor (.47) and negatively on the EC factor (−.66). Lastly, exploration of objects loaded positively onto the PE factor (.62) and negatively onto the NE factor (−.54).

3.2. Bivariate correlations

A multi-trait multi-method approach to examining the validity of experimenter ratings was adopted and the intercorrelations across composite traits (PE, NE, and EC) and method of assessment (experimenter ratings, objective coding, and parent report) are depicted in Table 2. The experimenter rating composite scores were comprised of the average score of the following items: PE (overall positive affect, energy, exuberance, social engagement with child tester, low shyness, enthusiasm toward tasks, initiative with tasks, anticipatory positive affect, exploration of objects, interest in test materials and stimuli, social engagement with parent); NE (prone to sadness, fear, and overall negative affect); EC (low cooperation with child tester, low attention to tasks, low cooperation with parent, low adaptation to change in test materials, impulsivity, low contentment, hyperactivity, low persistence in attempting to complete tasks, prone to anger/irritability, frustra-

Table 2
Bivariate associations between temperament traits.

	Traits	Experimenter ratings			Objective coding			Parent report		
		PE_1	NE_1	EC_1	PE_2	NE_2	EC_2	PE_3	NE_3	EC_3
Experimenter ratings	PE_1	(.90)								
	NE_1	−.46**	(.71)							
	EC_1	.13	−.51**	(.89)						
Objective coding	PE_2	.69**	−.26**	.10	(.87)					
	NE_2	−.06	.53**	−.52**	.07	(.63)				
	EC_2	−.31**	−.19*	.56**	−.38**	−.36**	(.81)			
Parent report	PE_3	.36**	−.06	−.23*	.27**	.17	−.26**	(.73/.76)		
	NE_3	−.12	.13	.13	−.12	.01	−.03	−.16	(.81/.78)	
	EC_3	−.10	−.05	.31**	−.12	−.27**	.35**	−.39**	−.17	(.84/.89)

Note. PE = Positive Emotionality, EC = Effortful Control, NE = Negative Emotionality. The values in parentheses are Cronbach's alphas.
** $p < .01$.
* $p < .05$.

tion with inability to complete tasks). Objective coding composite scores were comprised of the average of the following behavior codes: PE (anticipatory positive affect, sociability, interest/engagement, overall positive affect); NE (fear, sadness, anger); EC (compliance, low activity level, low impulsivity). Parent report composite scores for mothers and fathers were calculated by averaging the following CBQ scales: Extraversion/Surgency (Impulsivity, High Intensity Pleasure, Activity Level, and low Shyness); NE (Discomfort, Sadness, Fear, Anger/Frustration, and low Soothability); EC (Low Intensity Pleasure, Smiling/Laughter, Inhibitory Control, Perceptual Sensitivity, and Attentional Control).

Experimenter rating factor score intercorrelations were moderate-to-large. The PE and NE composites were negatively associated (−.46), as were NE and EC (−.51). Experimenter ratings of PE and EC were moderately positively correlated (.13). The data were scrutinized for evidence of convergent validity (indicated by the magnitude of correlations across unique methods of assessing the same trait) and discriminant validity (unique methods of assessing different traits do not correlate with one another). Convergent validity was evidenced by large intercorrelations between the three higher-order temperament dimensions assessed via experimenter report and objective coding of child traits (PE = .69, NE = .53, EC = .56), as well as modest-to-moderate intercorrelations between experimenter ratings and parent report (PE = .36, NE = .13, EC = .31), and objective coding and parent report (PE = .27, NE = .01, EC = .35). Discriminant validity was also generally observed, such that unique methods of assessing different temperament traits were either uncorrelated, or modestly correlated with one another. However, large intercorrelations were observed between experimenter ratings of EC and objective coding ratings of NE (−.52). The strong relationship between EC and NE may be attributable to high degree of overlap between the experimenter rating items that tapped dimensions of both of these higher-order traits. Children rated by experimenters as exhibiting low EC were also likely to be rated low on items assessing cooperation and adaptability to change. We also observed significant associations between EC and PE, such that objective coding ratings of EC and PE were moderately negatively correlated (−.38), as were parent report of PE and EC (−.39) on the CBQ.

4. Discussion

The current study examined the structure and convergent and discriminant validity of child traits assessed by experimenter ratings of child behavior during laboratory tasks. As hypothesized, a three-factor solution (PE, NE, and EC) fit the data best. The results of this examination are consistent with other investigations, in which similar three-factor models of child temperament were revealed (Ahadi et al., 1993; Casalin et al., 2012; Rothbart et al., 2001).

They differ somewhat from the findings of Dyson et al. (2011), who recovered five factors from objective coding of child behavior in response to lab tasks very similar to the ones described in this report. However, our results regarding the structure of experimenter ratings are similar to an investigation of personality in early childhood (Wilson, Schalet, Hicks, & Zucker, 2013) using a different method, the California Child Q-Set (CCQ; Block & Block, 1980), which was completed by trained examiners following administration of neurocognitive tasks to a sample of 3–5 year old children (Zucker et al., 2000) study. Hierarchical cluster analyses of the examiner ratings revealed two broad clusters: Adaptive Socialization (e.g., emotional stability, compliance, intelligence) and Anxious Introversion (e.g., emotional and behavioral introversion). While two clusters (not three) were identified, the Adaptive Socialization cluster resembles the PE and EC factors (e.g., interest, compliance, attention) identified in the current sample, and the Anxious Introversion cluster resembles the NE factor (e.g., fear, sadness).

Convergent validity between experimenter ratings and objective coding of child behavior was also demonstrated via modest-to-large convergent intercorrelations between the experimenter rating factor scores, objective coding of similar traits and parent-report on the CBQ, providing additional support for the three-factor solution. Experimenter ratings also exhibited good discriminant validity in that the three experimenter rating factor scores were either uncorrelated with or less strongly correlated with objective coding of traits marking different factors and parent report of distinct traits. Thus, in terms of structural and convergent and discriminant validity, our results provide support for the use of experimenter ratings of child traits based on observations of child behavior in response to structured laboratory tasks. The major advantage of this approach is that experimenter ratings can be collected and scored in a fraction of the time it takes to complete objective coding of child behavior.

The current study contributed to the existing literature in several ways. First, we explored the structure of child temperament traits via experimenter ratings of child behavior, in contrast to the oft relied upon informant (i.e., parent) report. Laboratory approaches, as opposed to questionnaire methods, are more cumbersome to utilize, which prevents their broader use in the literature. However, behavioral ratings have several strengths, including providing a structured examination of and more direct means of quantifying individual differences in children's traits. Additionally, as the convergence between different methods of assessing child behavior tends to be modest-to-moderate at best, continued investigation of the areas of convergence and divergence across methods is necessary to uncover the structure of child temperament and the functional significance of these traits.

Second, the results support the validity of experimenter ratings of child behavior as a means of indexing child temperament traits. Experimenter ratings share commonalities with both objective coding of child behavior and parent-report. Similar to objective coding methods, experimenter ratings are completed by people who have no previous experience with or relationship with the child. However, similar to parent-report, experimenter ratings culminate in aggregate scores across a myriad of child temperament traits (observed across the length of the lab visit), and the reporter (the experimenter) has an interactive role with the child as they engage with him/her during the laboratory tasks. Furthermore, experimenter ratings provide an interesting hybrid between behavioral observation and informant-report and are an efficient and economical means of evaluation, as they can be collected immediately following the laboratory visit and readily scored. Thus, the results provide support for the use of experimenter ratings as a substitute for or supplement to objective coding measures.

Future research should strive to build on the limitations of the current study. First, this study is cross-sectional and therefore unable to explore how the structure of child temperament may change with age. Future studies should consider conducting longitudinal examinations over the course of childhood to address this gap in the literature. Additionally, the current study only assessed the structure of child temperament in children ages 3–7 years. Less is known about the nature of temperament in younger children and infants. Also, experimenter reports were conducted by a single experimenter following each visit. As such, we were unable to examine inter-experimenter agreement for these ratings. Such data could lend additional support for the validity of experimenter ratings. Finally, although factor analysis is useful for summarizing the covariance in measured variables, the accuracy of the recovered structure depends upon the nature of those variables, the

sample in which they were measured, and the clarity of the structure relative to other alternatives. Furthermore, the existence of a particular structure does not reveal the underlying developmental processes that produced the observed covariance structure. Evidence of differential influences upon or outcomes related to these three traits generated by longitudinal studies would provide important evidence for the validity of structural analyses such as those we report. In conclusion, the results support a three-factor model of temperament (PE, NE, and EC), with convergent and discriminant validity between experimenter ratings and objective coding of child behavior providing additional support for this model. Furthermore, experimenter ratings completed immediately following a laboratory visit provide a convenient and economical proxy for objective coding of the laboratory tasks.

References

Ahadi, S. A., Rothbart, M. K., & Ye, R. (1993). Children's temperament in the U.S. and China: Similarities and differences. *European Journal of Personality, 7*, 359–378. doi: http://dx.doi.org/10.1002/per.2410070506.

Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. *Proceedings of Third Berkley Symposium on Mathematical Statistics and Probability, 5*, 111–150.

Block, J., & Block, J. H. (1980). *The California Child Q-set*. Palo Alto, CA: Consulting Psychologists Press. Original work published in 1969.

Buss, A. H., & Plomin, R. (1984). *Temperament: Early developing personality traits*. Hillsdale, NJ: Erlbaum.

Carlson, S. M., & Wang, T. S. (2007). Inhibitory control and emotion regulation in preschool children. *Cognitive Development, 22*, 489–510. http://dx.doi.org/10.1016/j.cogdev.2007.08.002.

Casalin, S., Luyten, P., Vliegen, N., & Meurs, P. (2012). The structure and stability of temperament from infancy to toddlerhood: A one-year prospective study. *Infant Behavior and Development, 35*, 94–108. http://dx.doi.org/10.1016/j.infbeh.2011.08.004.

Comrey, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology, 56*, 754–761. doi: http://dx.doi.org/10.1037//0022-006X.56.5.754.

De Pauw, S., & Mervielde, I. (2010). Temperament, personality and developmental psychopathology: A review based on the conceptual dimensions underlying childhood traits. *Child Psychiatry and Human Development, 41*, 313–329.

De Pauw, S. S. W., Mervielde, I., & Van Leeuwen, K. G. (2009). How are traits related to problem behavior in preschoolers? Similarities and contrasts between temperament and personality. *Journal of Abnormal Child Psychology, 37*, 309–325. http://dx.doi.org/10.1007/s10802-008-9290-0.

Dennis, T. A., Brotman Miller, L., Huang, K., & Kiely Gouley, K. (2007). Effortful control, social competence, and adjustment problems in children at risk for psychopathology. *Journal of Clinical Child and Adolescent Psychology, 36*, 442–454. http://dx.doi.org/10.1080/15374410701448513.

Digman, J. M., & Shmelyov, A. G. (1996). The structure of temperament and personality in Russian Children. *Journal of Personality and Social Psychology, 71*, 341–351.

Durbin, C. E. (2010). Validity of young children's self-reports of their emotion in response to structured laboratory tasks. *Emotion, 10*, 519–535. http://dx.doi.org/10.1037/a0019008.

Durbin, C. E., Hayden, E. P., Klein, D. K., & Olino, T. M. (2007). Stability of laboratory-assessed temperamental emotionality traits from ages 3 to 7. *Emotion, 7*, 188–199. http://dx.doi.org/10.1037/1528-3542.7.2.388.

Durbin, C. E., Klein, D. N., Hayden, E. P., Buckley, M. E., & Moerk, K. C. (2005). Temperamental emotionality in preschoolers and parental mood disorders. *Journal of Abnormal Psychology, 114*, 28–37.

Durbin, C. E., & Wilson, S. (2012). Convergent validity of and bias in parent reports of child emotion. *Psychological Assessment, 24*, 647–660.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary Test* (3rd ed.). Circle Pines, Minnesota: American Guidance Service.

Dyson, M. W., Olino, T. M., Durbin, C. E., Hill Goldsmith, H., & Klein, D. N. (2011). The structure of temperament in preschoolers: A two-stage factor analytic approach. *Emotion, 12*, 44–57. http://dx.doi.org/10.1037/a0025023.

Gagne, J. R., Van Hulle, C. A., Aksan, N., Essex, M. J., & Goldsmith, H. H. (2011). Deriving childhood temperament measures from emotion-eliciting behavioral episodes: Scale construction and initial validation. *Psychological Assessment, 23*, 337–353. http://dx.doi.org/10.1037/a0021746.

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

Goldsmith, H. H., Reilly, J., Lemery, K. S., Longley, S., & Prescott, A. (1995). *Laboratory temperament assessment battery: Preschool version*. Unpublished manuscript.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Halverson, C. F., Havill, V. L., Deal, J., Baker, S. R., Victor, J. B., Pavlopoulos, V., et al. (2003). Personality structure as derived from parental ratings of free descriptions of children: The inventory of child individual differences. *Journal of Personality, 71*, 995–1026. doi: http://dx.doi.org/10.1111/1467-6494.7106005.

Izard, C., Dougherty, L., & Hembree, E. (1983). *A system for identifying affect expressions by holistic judgments (AFFEX)*. Newark, DE: Instructional Resources Center, University of Delaware.

Kagan, J., Snidman, N., McManis, M., Woodward, S., & Hardway, C. (2002). One measure, one meaning: Multiple measures, clearer meaning. *Development and Psychopathology, 14*, 463–475.

Kochanska, G., Murray, K., Jacques, T. Y., Koenig, A. L., & Vandegeest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development, 67*, 490–507. doi: http://dx.doi.org/10.2307/1131828.

Kotelnikova, Y., Olino, T. M., Mackrell, S. V. M., Jordan, P. L., & Hayden, E. P. (2013). Structure of observed temperament in middle childhood. *Journal of Research in Personality, 47*, 524–532.

Majdandzic, M., & van den Boom, D. C. (2007). Multimethod longitudinal assessment of temperament in early childhood. *Journal of Personality, 75*, 121–167. http://dx.doi.org/10.1111/j.1467-6494.2006.00435.x.

Olino, T. M., Durbin, C. E., Klein, D. N., Hayden, E. P., & Dyson, M. W. (2013). Gender differences in young children's temperament traits: A multi-method analysis. *Journal of Personality, 81*, 119–129.

Presley, R., & Martin, R. P. (1994). Toward a structure of preschool temperament: Factor structure of the temperament assessment battery for children. *Journal of Personality, 62*, 415–448.

Rothbart, M. K. (1981). Measurement of temperament in infancy. *Child Development, 52*, 569–578. doi: http://dx.doi.org/10.1111/j.1467-8624.1981.tb03082.x.

Rothbart, M. K. (2007). Temperament, development, and personality. *Current Directions in Psychological Science, 16*, 207–212. http://dx.doi.org/10.1111/j.1467-8721.2007.00505.x.

Rothbart, M. K., & Derryberry, D. (1981). Development of individual differences in temperament. In M. E. Lamb & A. L. Brown (Eds.). *Advances in developmental psychology* (Vol. 1, pp. 37–86). Hillsdale, NJ: Erlbaum.

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., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at 3–7 years: The children's behavior questionnaire. *Child Development, 72*, 1394–1408. doi: http://dx.doi.org/10.1111/1467-8624.00355.

Shiner, R. L., Buss, K. A., McClowry, S. G., Putnam, S. P., Saudino, K. J., & Zenter, M. (2012). Temperament research on the twenty-fifth anniversary of Goldsmith et al. (1987). *Child Development Perspectives, 6*, 436–444. http://dx.doi.org/10.1111/j.1750-8606.2012.00254.x.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420–428.

Stifter, C. A., Willoughby, M. T., & Towe-Goodman, N. (2008). Agree or agree to disagree? Assessing the convergence between parents and observers on infant temperament. *Infant and Child Development, 17*, 407–426.

Streiner, D. L. (1994). Figuring out factors: The use and misuse of factor analysis. *Canadian Journal of Psychiatry, 39*, 135–140.

Tellegan, A. (1985). Structure of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In A. H. Tuma & J. D. Maser (Eds.), *Anxiety and the anxiety disorders* (pp. 681–706). Hillsdale, NJ: Lawrence Erlbaum Associates Inc..

Thomas, A., & Chess, S. (1977). *Temperament and development*. Oxford, England: Brunner/Mazel.

Willoughby, M. T., Wirth, R. J., & Blair, C. B. (2012). Executive function in early childhood: Longitudinal measurement invariance and developmental change. *Psychological Assessment, 24*, 418–431. http://dx.doi.org/10.1037/a0025779.

Wilson, S., & Durbin, C. E. (2012). Dyadic parent-child interaction during early childhood: Contributions of parental and child personality traits. *Journal of Personality, 80*, 1313–1338.

Wilson, S., Schalet, B. D., Hicks, B. M., & Zucker, R. A. (2013). Identifying early childhood personality dimensions using the California Child Q-set and prospective associations with behavioral and psychosocial development. *Journal of Research in Personality, 47*, 339–350. doi: http://www.dx.doi.org/10.1016/j.jrp.2013.02.010.

Zucker, R. A., Fitzgerald, H. E., Refior, S. K., Puttler, L. I., Pallas, D. M., & Ellis, D. A. (2000). The clinical and social ecology of childhood for children of alcoholics: Description of a study and implications for a differentiated social policy. In H. E. Fitzgerald, B. B. Lester, & B. S. Zuckerman (Eds.), *Children of addiction: Research, health, and policy issues* (pp. 1–30). New York: Garland Press.