

Ecological Validity of Laboratory Assessments of Child Temperament: Evidence From Parent Perspectives

Sharon L. Lo, Lisa N. Vroman, and C. Emily Durbin
Michigan State University

The structure, stability, and validity of child temperament traits have primarily been examined with parent questionnaire methods, but laboratory methods represent an important complement. However, the novel setting and contrived scenarios of laboratory methods and their low convergence with parent questionnaire methods have led some to question their ecological validity. We tested this assumption by employing parents as sources of information regarding the ecological validity of laboratory assessments of child temperament. Parents observed their child participating in 10 different laboratory tasks and reported on the typicality of their child's behavior. The results suggested parents considered their child's responses during the laboratory tasks as highly typical representations of their child's behavior outside of the lab, supporting the ecological validity of trait-relevant behavior elicited with laboratory tasks.

Keywords: child temperament, assessment, parent report, ecological validity

Individual differences in emotional reactivity and regulation, known as temperament, are thought to reflect some of the earliest appearing dispositions that have both rank-order stability and continuity with later outcomes. Evidence suggests that temperament traits are in fact reasonably stable over considerable developmental intervals (Caspi & Roberts, 2005; Goldsmith et al., 1987) and that they are predictors of concurrent and future psychosocial adjustment (e.g., Caspi, Moffitt, Newman, & Silva, 1996; Shiner, 2000). Active foci of current research on child temperament include issues of developmental change in the manifestation, structure, and mean levels of traits across childhood and adolescence, as well as sources of influence on individual differences in temperament traits. However, the bulk of empirical research testing models of child temperament has been derived from a single approach to assessing temperament: parent questionnaire methods. A literature search for peer-reviewed publications from 1980 to 2014 listed under the term *child temperament* in the PsycINFO and PsycARTICLES databases yielded a total of 4,291 papers. Within these publications, the search term *laboratory assessment* yielded 234 papers. A majority (89%) of these papers were published after the year 2000. Of the 156 papers that were available and provided a specific number of laboratory tasks used in the study, 103 studies used 5 or fewer tasks (66%) and 53 studies used 6 or more tasks (34%). Thus, approximately 5% of publications on child temperament (published between 1980 and 2014) included laboratory methods (single task, several tasks, or full battery), with even fewer using an extensive laboratory battery. Overreliance on a single method reduces confidence in the generalizability of sub-

stantive findings, raising the possibility that findings are unique to a particular method or driven by shared method variance across predictor and outcome (e.g., as in studies in which child traits and adjustment are both measured via parent questionnaire). Just as it is important not to overrely on one method of assessment, it is also critical to assess the ecological validity of other various assessment methods such as laboratory tasks. The use of multiple methodological approaches also increases the richness of our understanding of temperament constructs, as different methods may contribute incremental information or novel insights regarding the nature of traits and their associations with external criteria.

Researchers have long debated the “optimal” strategy for assessing child temperament, an argument that orients the field toward a somewhat simplistic enumeration and defense of the relative advantages of different approaches and away from clarifying the content domain of temperament and exploring the mechanisms that underlie traits and their development. For example, parent report questionnaires are praised for their economy and in particular their ecological validity, or ability to represent individual differences in behaviors and traits as these are revealed in a naturalistic setting. By contrast, methods that are less economical and more time intensive, such as laboratory approaches, are viewed as providing more direct evidence for individual differences in traits than parent reports (Durbin, 2010), but they are criticized in that the novel setting and contrived scenarios of the laboratory are believed to reduce the ecological validity of this approach (e.g., Dhimi, Hertwig, & Hoffrage, 2004; Majdandžić & Van Den Boom, 2007). Such concerns may reduce the likelihood that a researcher will use laboratory approaches, potentially contributing to the relatively low representation of lab methods in the literature on child temperament. These arguments pit different approaches against one another that choosing the “best” approach requires privileging either rigor and standardization (laboratory tasks) or economy and ecological validity (parent report). However, rather than submitting to such a choice, it may be more important to encourage the use of multiple methods of assessing

This article was published Online First October 20, 2014.

Sharon L. Lo, Lisa N. Vroman, and C. Emily Durbin, Department of Psychology, Michigan State University.

Correspondence concerning this article should be addressed to Sharon L. Lo, Department of Psychology, Michigan State University, 40A Psychology Building, East Lansing, MI 48824. E-mail: losharon@msu.edu

child temperament. This approach builds upon the strengths of each method but also allows one to use different approaches to inform interpretations of data generated by any particular method.

Parent Report Method of Assessing Temperament

Parent report measures of temperament have provided a rich literature on continuity between child and adult personality (e.g., Rothbart, Ahadi, Hershey, & Fisher, 2001) and on the rank-order stability of traits from infancy to middle childhood (e.g., Lemery, Goldsmith, Klinnert, & Mrazek, 1999; Pedlow, Sanson, Prior, & Oberklaid, 1993; Roberts & DelVecchio, 2000). This is by far the most common approach to assessing child temperament and is therefore responsible for most of what we know about child traits and their correlates. Parent questionnaires are cost effective, thus facilitating their use with large samples, and they use source informants who are privy to the child's entire developmental history and who are also most directly invested in understanding the child as a person. However, relying primarily on parent report methods to assess temperament has a number of limitations. For example, parent report methods have low convergence with naturalistic observations and teacher reports (e.g., Seifer, Sameroff, Barrett, & Krafchuk, 1994), and one study showed they have weaker predictive validity for later psychological functioning (as reported by the child) than do teacher reports (Mesman & Koot, 2000).

Finally, there is evidence that parent reports are influenced/biased by parental characteristics (De Los Reyes & Kazdin, 2005), particularly maternal depression and anxiety (Boyle & Pickles, 1997; Chi & Hinshaw, 2002; Gartstein, Bridgett, Dishion, & Kaufman, 2009; Youngstrom, Izard, Ackerman, 1999) but also parental and familial stress (Foley, Rutter, Angold, & Pickles, 2005; Youngstrom, Loeber, & Stouthamer-Loeber, 2000), low socioeconomic status (Duhig, Renk, Epstein, & Phares, 2000), maternal alcoholism (Foley et al., 2005), and marital discord (Christensen, Margolin, & Sullaway, 1992; Webster-Stratton & Hammond, 1988). Moreover, bias is evident even in conditions intended to maximize convergence. For example, Durbin and Wilson (2012) had mothers and naive observers watch the same brief samples of child behavior, then immediately report on the child's emotions; bias attributable to maternal depression and anxiety was evident even in these reports. Given replicated evidence of bias in parental reports, some have questioned whether the stability of temperament traits assessed via parent report is inflated by the influence of parents' own stable traits over time (Bates, 1994) or perhaps by parents' efforts to maintain a consistent view of their child (Kagan, 1998).

Laboratory Method of Assessing Temperament

Although laboratory assessment methods involve a greater time investment and additional expense compared to questionnaires, they also offer a number of unique advantages relative to parent report methods. For example, the standardized stimuli and procedures used in laboratory measures are designed to elicit sharper behavioral evidence of individual differences in temperament traits of interest, allowing for a more direct assessment of the full range of individual differences in these

traits. Given the experimental control over the stimuli employed in standardized lab tasks, laboratory measures can also be designed to elicit behaviors that are expressed at a lower base rate in naturalistic contexts, such as fear. The objective coding criteria used to assess child behavior during lab tasks contrasts with the known influence of parental characteristics on their ratings of child behavior. The validity of this approach is supported by evidence indicating laboratory methods have predictive validity for both normal developmental outcomes, such as moral development (Kochanska, Murray, Jacques, Koenig, & Vandegest, 1996) and aspects of the parent-child relationship (Kochanska, Aksan, & Carlson, 2005), and risk for problematic outcomes, including internalizing disorders (Dougherty, Klein, Durbin, Hayden, & Olino, 2010; Durbin, Klein, Hayden, Buckley, & Moerk, 2005; Hayden, Klein, Durbin, & Olino, 2006).

One of the primary critiques of laboratory methods is their low-to-moderate convergence with parent report. These data are often used to support the claim that laboratory assessments lack ecological validity (e.g., Dhami et al., 2004; Majdandžić & Van Den Boom, 2007). However, other researchers have highlighted that high congruence between multiple methods should not be expected because every method has distinct advantages and limitations (Rothbart & Goldsmith, 1985); instead, the recommended approach for studying temperament would be using multiple methods of assessment (Kagan, Snidman, McManis, Woodward, & Hardway, 2002). Methods that differ considerably in their relative strengths (e.g., parent report and laboratory tasks) may be maximally and mutually informative, as they test the limits of trait coherence and provide a means of testing assumptions inherent in the other method.

Current Study

The present study explored the ecological validity of laboratory assessments of child temperament by drawing upon parents' knowledge of their child's behavior. Parents observed their child participating in structured laboratory tasks and following each task reported on the degree to which the child's responses were typical of his or her behavior outside of the lab (i.e., in more naturalistic settings). To our knowledge, no previous studies have explored the ecological validity of laboratory tasks in this way. We used these reports to test several hypotheses. First, we quantified the ecological validity of the laboratory approach by examining mean levels of and variability in parent-reported typicality of child behavior in response to lab tasks. Second, we compared parent ratings of child typicality across lab tasks designed to elicit positive emotions versus those intended to elicit negative emotions. On the basis of the literature suggesting that parents have a tendency to rate positive attributes of their child as more consistently true than negative ones (e.g., Gretarsson & Gelfand, 1988), we hypothesized that, on average, parent typicality ratings would be higher for lab tasks designed to elicit positive emotions. Finally, we examined whether individual differences in any child traits were associated with parent reports of higher or lower typicality by examining bivariate and within-subject associations between parent typicality ratings and objective coding of child emotions and behaviors.

Method

Participants and Procedure

Participants were 206 children between the ages of 3 and 7 years recruited from the greater Chicago area for a study of child temperament. Children who did not have any significant medical conditions or developmental disabilities and lived with at least one English-speaking parent were eligible for participation in the study. The final sample of children had a mean age of 56.4 months ($SD = 12.0$; range = 36–83), and 48.1% were girls. Mothers were between the ages of 23 and 49 years ($M = 36.9$, $SD = 4.8$), and fathers were between the ages of 23 and 57 years ($M = 38.8$, $SD = 5.8$). Data on race and 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.

Child participants visited the laboratory with their mother ($n = 191$) or father ($n = 14$) for a 2-hour assessment consisting of tasks designed to elicit discrete emotions and behaviors indicative of temperament traits. Children were also administered the Peabody Picture Vocabulary Test, Second Edition (PPVT; Dunn & Dunn, 1997), at the beginning of the lab visit to assess their level of receptive language skills ($M = 106.62$, $SD = 15.07$). At the end of the lab visit, the parent was given a battery of questionnaires to complete and return by mail.

Laboratory Assessment of Temperament

Child temperament was assessed with a 2-hour battery of 10 structured tasks composed of episodes from the Laboratory Temperament Assessment Battery–Preschool Version (Lab-TAB; Goldsmith, Reilly, Lemery, Longley, & Prescott, 1995) or tasks newly developed for this study (i.e., Exploring New Objects, Making a T-Shirt, Disappointing Toy, and Dress-Up). The structured tasks were designed primarily to elicit behaviors and emotional responses indicative of individual differences in traits related to positive emotionality (PE; positive mood, sociability, activity level, engagement) and negative emotionality (NE; fear-, sadness-, and anger-proneness). Prior to the assessment, the parent was instructed to respond neutrally to the child's advances and limit his or her interaction with the child. The parent accompanied the child into the assessment room for all but three tasks (Stranger Approach, Pop-Up Snakes, and Box Empty), for which the parent was directed to the camera room. There, the child was observed through a one-way mirror. Between tasks, a short play break (approximately 2–3 minutes) was used to allow children to return to a baseline affective state. Episodes were ordered such that those eliciting similar emotions were not presented consecutively. Lab tasks are described below in the order that the episodes were conducted during the assessment. All lab episodes were videotaped for later coding.

Exploring New Objects (fear, happiness, engagement, activity level). The child was instructed to explore a room alone for 5 minutes. The room contained novel objects such as a large plastic

skull hidden beneath a red cape, a remote-controlled spider, and an animal crate filled with furry toy mice.

Making a T-Shirt (engagement, happiness). The experimenter presented a blank white T-shirt to the child and provided instruction on how to use fabric markers to decorate the shirt. The child was then allowed to independently decorate the T-shirt for 2 minutes.

Disappointing Toy (anger, sadness). The child was presented with pictures of appealing toys (e.g., puppets, remote-controlled race cars) and an unappealing toy (plastic watering can) and was asked to select his or her favorite toy from the pictures. The experimenter returned with the unappealing toy. After 2 minutes, the experimenter returned with the toy the child initially chose, and they played together with it.

Stranger Approach (fear, sociability). The child was told to wait alone in the room for a moment. After 20 seconds, a male research assistant who the child had not seen before entered the room and had a brief interaction with the child based on a neutral scripted conversation.

Dress-Up (engagement, happiness, sociability). The child and experimenter played with costumes (e.g., fireman's jacket and hat, doctor's jacket and stethoscope) for 2 minutes. The experimenter then took a Polaroid photo of the child in his or her costume.

Transparent Box (anger, sadness, engagement, initiative). The experimenter presented two appealing toys (e.g., trucks, Barbies, stuffed animals), and the child picked a favorite to lock inside a transparent box. The child was left with a set of nonfunctional keys for 3 minutes. The experimenter then returned, explained that she made a mistake, and provided the child the right set of keys to unlock the box and play with the toy.

Popping Bubbles (activity level, happiness, sociability, impulsivity). The experimenter and child took turns playing with a bubble-making toy for 3 minutes. The experimenter instructed the child to try popping the bubbles with different body parts (e.g., hands, feet).

Green Circles (anger, sadness). The experimenter asked the child to draw a perfect green circle. The experimenter mildly criticized the child's green circle for little imperfections, such as its size or shape, and repeatedly asked the child to draw another green circle. After 2 minutes, the experimenter commented positively about all of the child's drawings.

Pop-Up Snakes (happiness, surprise, sociability, engagement). The experimenter pretended to struggle with opening a can of chips for 10 seconds and asked the child for help. The child opened the can of chips to find two coiled-spring snakes that fly out of the can. The child was then encouraged to also surprise his or her parent with the pop-up snakes.

Box Empty (anger, sadness). The child was given a gift-wrapped empty box under the impression there was an appealing toy inside. After a period of 2 minutes when the child was left alone to discover the gift bag was empty, the experimenter returned with several toys for the child to take home, explaining that she had forgotten to place them in the gift box.

Coding of Child Temperament

Laboratory episodes were coded with a global system (Durbin et al., 2005). Selection of traits was based on previous literature on the

structure of temperament as assessed via laboratory tasks (see Dyson, Olino, Durbin, Goldsmith, & Klein, 2012). PE (happiness), NE (anger, fear, and sadness), and surprise were assessed by coding and counting discrete instances of facial, vocal, and bodily expressions of each emotion. For all 10 episodes, time-stamp recorded instances of discrete emotions were coded as low, moderate, or high intensity. Intensity ratings were summed across facial, vocal, and bodily ratings such that each emotion had three total intensity scores (e.g., low fear, moderate fear, high fear) for each episode. A weighted average of these scores was created by converting intensity level to a 3-point scale (1 = *low intensity*, 2 = *moderate intensity*, 3 = *high intensity*), to yield weighted composite scores for happiness, surprise, anger, fear, and sadness. Reliability of weighted average coding scores was indexed on a subsample of participants (15%, $N = 27$). Interrater reliability of these scores was calculated with intraclass correlations (case 1, Shrout & Fleiss, 1979). Intraclass correlation coefficients (ICCs) for total weighted expression counts were as follows: PE (.90), fear (.66), sadness (.79), anger (.81), and surprise (.65).

Coders rated several other traits by watching an entire lab task and assigning a single rating for each variable based on all relevant behaviors that were observed during the task. The following variables were rated on a 4-point Likert scale (0 = *low*, 1 = *moderate*, 2 = *moderate-to-high*, and 3 = *very high*): Interest/engagement was based on how engaged and persistent the child was during the task. Activity level was based on the child's overall movement around the room and vigor in manipulating stimuli. Anticipatory PE was based on the child's positive behavioral and emotional response in anticipation of a positive event. Initiative was based on the child's degree of passivity or assertiveness in interactions with the experimenter or parent. Sociability referred to the child's display of social referencing and attempts to engage with the experimenter or parent. Compliance ratings considered the severity of the child's deliberate unwillingness to comply with the experimenter's or parent's demands or suggestions. Impulsivity was based on the child's tendency to act or respond without reflection or hesitation. Global behavior ratings were averaged across all 10 episodes to yield composite scores of interest, activity level, anticipatory PE, initiative, sociability, compliance, and impulsivity. ICCs for the child temperament subtraits were generally in the adequate-to-excellent range (Mitchell, 1979), ranging from .65 (engagement) to .94 (activity). Scores for each dimension (PE, fear, sadness, anger, interest, activity level, anticipatory PE, initiative, sociability, compliance, and impulsivity), averaged across all 10 tasks, were subjected to principal-components analysis with a varimax rotation (Wilson & Durbin, 2012). Results revealed three higher order temperament dimensions: PE (PE, anticipatory PE, sociability, engagement), NE (fear, sadness, anger), and effortful control (EC; compliance, low activity level, low impulsivity). We have previously shown that traits assessed with this objective coding scheme show stability across ages 3 to 7 (Durbin, Hayden, Klein, & Olino, 2007) and that they are associated with familial risk for psychopathology (Durbin et al., 2005; Olino, Klein, Dyson, Rose, & Durbin, 2010) and with children's self-reported emotions (Durbin, 2010).

Parental Report of Child Emotion and Typicality

The parent was instructed to rate the child's emotions and typicality of the child's behavior in response to each lab task. The experimenter described the rating form to the parent and instructed the parent to base ratings on his or her observations during the task. After each task,

the parent rated the frequency and intensity with which the child experienced the following emotions: happiness, contentment, surprise, fear, nervousness, sadness, and anger or irritability. Frequency was assessed with the following item: "In your opinion, how frequently (if at all) did your child experience the following emotions during the task?" Frequency was rated on the following scale: 1 = *never felt the emotion*, 2 = *less than half the time*, 3 = *about half the time*, 4 = *more than half the time*, 5 = *almost all the time*. Intensity was assessed with the following item: "In your opinion, how intensely did your child experience the following emotions during this task? Please rate how strongly the child felt the emotion." Intensity was rated on the following scale: 1 = *emotion never present*, 2 = *low intensity/not very strongly*, 3 = *moderate intensity/somewhat strongly*, 4 = *high intensity/quite strongly*, 5 = *very high intensity/very strongly*. They also rated the sources of information they drew upon in making these ratings on the following scale: 1 = *did not use at all; no influence on my decision*, 2 = *used a little bit; some influence on my decision*, 3 = *used a lot; strong influence on my decision*. The following sources were included: facial expressions; bodily movements; what your child said; how physically active your child was; what you thought your child was thinking; how your child has reacted in the past to similar situations; how you thought most children would react; and how interested in the task you thought your child was. These ratings are further described in a separate report (Durbin & Wilson, 2012) and are not considered further herein.

Parent ratings also included one item asking how typical the child's behavior during the episode was of his or her usual behavior: "How typical of your child's usual behavior was his/her behavior during this task?" Typicality was rated as follows: 1 = *not at all typical—my child almost never acts like this*, 2 = *somewhat typical—my child acts like this on some occasions*, 3 = *moderately typical—my child acts like this fairly often*, 4 = *highly typical—my child regularly acts like this*, 5 = *extremely typical—my child almost always acts like this*. We examined parent typicality ratings at several levels, including for individual lab tasks, aggregated across positive and across negative tasks, and averaged across all 10 lab tasks. Parent-reported typicality ratings exhibited high internal consistency reliability across all 10 tasks ($\alpha = .82$).

Data Analysis

We report on (a) mean-level differences in parent-reported typicality of child behavior across lab tasks designed to elicit positive and negative emotions, (b) bivariate and within-subject correlations between parent-reported typicality and objective coding of child emotions and behaviors, (c) one-way between-subjects analysis of variances (ANOVAs) to compare individual differences in parent-reported child emotion between low parent typicality and high parent typicality groups, and (d) ANOVAs to compare individual differences in objectively coded child emotion between low parent-rated typicality and high parent-rated typicality groups.

Results

Parent-Reported Typicality Ratings Across the Laboratory Battery

Parent-reported typicality ratings had a mean and mode of 4.00, $SD = 0.93$ (recall that 4 = *highly typical*), suggesting that parents considered their child's responses during lab tasks as highly typ-

ical representations of their child's behavior outside the lab (see Table 1). The modal typicality rating was a 5 (*extremely typical—my child almost always acts like this*) for two tasks and a 4 (*highly typical—my child regularly acts like this*) for the remaining 8 tasks. There were no mean-level differences in parent-reported typicality ratings between mothers and fathers, $t(202) = 1.10, p = .273$. As predicted, parent-reported typicality ratings for tasks designed to elicit positive emotions were significantly higher than in those designed to elicit negative emotions, $t(206) = 10.05, p < .001, d = 0.61$.¹ Objective coding of child emotions revealed that tasks designed to elicit positive emotions did have significantly higher average levels of coded PE than tasks designed to elicit negative emotions, $t(190) = 17.41, p < .001, d = 0.91$. Similarly, tasks designed to elicit negative emotions had significantly higher weighted average levels of coded fear, $t(189) = 11.91, p < .001, d = 1.10$; sadness, $t(189) = 14.68, p < .001, d = 1.38$; and anger, $t(189) = 14.93, p < .001, d = 1.41$, as compared to tasks designed to elicit positive emotions. This suggests that parents viewed their child's behavior in the lab setting as being more typical of his or her usual behavior during tasks eliciting higher levels of PE than during tasks that generally elicited higher levels of NE (see Table 1).

Associations Between Parent-Reported Typicality and Laboratory-Assessed Child Traits

Bivariate correlations. Bivariate correlations between parent-reported typicality (averaged across all 10 tasks) and objective coding of child traits revealed that PE, interest/engagement, initiative, and sociability were positively associated with parent ratings of average typicality (see Table 2). Conversely, fear, sadness, anger, activity, compliance, and impulsivity were not related to typicality ratings. Thus, behaviors indicative of NE and low EC were unrelated to typicality, whereas children who were coded as exhibiting more surgent and extraverted behavior were rated by the parent as exhibiting behavior in response to the lab tasks that was more typical of their usual behavior. These results are consistent with Funder's (1995) realistic accuracy model and data demonstrating that persons who are judged most accurately are those

Table 1
Parent-Reported Typicality Ratings Across Laboratory Tasks

Task type	Frequency counts (%)					Descriptives	
	1	2	3	4	5	<i>M</i>	<i>SD</i>
Positive						4.23	0.86
Dress-Up	2.40	3.40	12.0	36.1	43.8	4.18	0.94
Making a T-Shirt	0.00	7.70	16.8	40.4	33.7	4.02	0.91
Popping Bubbles	0.00	0.00	4.80	31.7	62.5	4.58	0.58
Pop-Up Snakes	0.05	4.80	14.4	40.4	36.5	4.11	0.87
Negative						3.85	0.94
Box Empty	1.90	8.70	16.8	44.2	27.9	3.88	0.98
Disappointing Toy	0.05	7.70	22.1	40.9	26.4	3.88	0.92
Exploring New Objects	0.00	7.20	15.4	43.3	31.7	4.02	0.89
Green Circles	1.00	7.20	29.3	38.0	16.8	3.68	0.89
Stranger Approach	0.05	11.5	17.3	37.5	25.0	3.82	0.99
Transparent Box	0.05	11.1	20.2	42.8	24.0	3.80	0.95

Note. Typicality rating scale: 1 = not at all typical; 2 = somewhat typical; 3 = moderately typical; 4 = highly typical; 5 = extremely typical.

Table 2
Bivariate and Within-Subject Correlations Between Objective Coding of Child Traits and Parent Reports of Typicality

Variable	Bivariate			Within-subject		
	<i>r</i>	<i>M</i>	<i>SD</i>	<i>r</i> ^a	<i>r</i> ^b	<i>r</i> ^c
Child trait						
PE	0.15*	40.2	19.6	0.23***	0.27***	0.09
Surprise	0.03	0.50	0.55	0.08	0.07	0.10
Anger	0.04	2.96	2.59	0.10*	0.34***	0.08
Fear	-0.03	2.64	3.18	0.25***	0.03	0.25***
Sadness	0.05	3.60	3.20	0.11**	0.15**	0.12*
Activity level	0.10	1.42	0.50	0.14	0.25***	0.12*
Anticipatory PE	0.14*	1.30	0.40	0.13	0.17*	0.09
Compliance	0.05	2.53	0.40	0.06	0.13	0.09
Impulsivity	0.02	0.40	0.36	0.08	0.09	0.09
Interest	0.26**	1.95	0.48	0.20*	0.33***	0.13
Initiative	0.23**	1.37	0.55	0.13	0.10	0.08
Sociability	0.21**	1.72	0.58	0.11	0.12	0.14**
Typicality ratings	1.00***	4.00	0.56			

Note. PE = positive emotionality.

^a Within-subject correlation between objective coding of child traits and parent reports of typicality averaged across all 10 tasks. ^b Within-subject correlation between objective coding of child traits and parent reports of typicality averaged across positive tasks only. ^c Within-subject correlation between objective coding of child traits and parent reports of typicality averaged across negative tasks only.

* $p < .05$. ** $p < .01$. *** $p < .001$.

whose traits are made more "available" by their saliency. This is particularly the case for those with high positive mood, sociability, and greater activity level.

Within-subject correlations. Within-subject correlations between parent-reported typicality and objective coding of child traits indicated that parent-typicality ratings were positively associated with within-child variation in PE across tasks ($r = .23, p < .001$), such that parents reported their child exhibited more typical behavior in those tasks in which she or he exhibited higher levels of PE. Higher typicality ratings were also associated with within-child variation in anger ($r = .10, p < .05$), fear ($r = .25, p < .001$), sadness ($r = .11, p < .01$), and interest ($r = .19, p < .05$) across tasks. Within-subjects correlations between typicality and objective coding of other traits (surprise, activity level, anticipatory PE, compliance, impulsivity, initiative, and sociability) were nonsignificant.

We also explored within-subject correlations between parent-reported typicality and objective coding of child traits separately for tasks designed to elicit positive emotions and those designed to elicit negative emotions. Results indicated that parent-typicality

¹ Parent typicality ratings for tasks designed to elicit positive emotions were higher for older children (6- to 7-year-olds) than for younger children (3- to 5-year-olds) even after accounting for the weighted average of coded expressions of PE that the child displayed across all lab tasks. This suggested that in lab tasks designed to elicit positive emotions, older children were rated as behaving more similarly to their usual behavior than were younger children. Parent typicality ratings did not differ in lab tasks designed to elicit negative emotion on the basis of the children's ages. Typicality ratings did not differ by parents' race/ethnicity or education level. Parent-reported personality traits related to PE, NE, and Constraint as measured by the Multidimensional Personality Questionnaire (Patrick, Curtin, & Tellegen, 2002) were also unrelated to parent-reported typicality.

ratings across positive tasks were positively associated with within-child variation in PE ($r = 0.27, p < .001$), anger ($r = .34, p < .001$), sadness ($r = .15, p < .010$), activity level ($r = .25, p < .001$), anticipatory PE ($r = .17, p < .050$), and interest ($r = .33, p < .001$) across those tasks. Higher typicality ratings across tasks designed to elicit negative emotions were positively associated with within-child variation in fear ($r = .25, p < .001$), sadness ($r = .12, p < .050$), activity level ($r = .12, p < .05$), and sociability ($r = .14, p < .010$) across negative tasks. Additional analyses were conducted to examine whether the association between higher typicality ratings and higher coded fear was specific to tasks designed to elicit fearful emotions. Results indicated that this relationship was observed in tasks designed to elicit fear ($r = .40, p < .001$) and was not observed in tasks designed to elicit frustration or anger ($r = .05, p = .94$). Taken together, our observations of within-child covariation between objective coding of child behavior in response to tasks and the parent's perceptions of the typicality of the child's behavior suggest that children who were more responsive to the lab tasks (i.e., more engaged, active, and exhibiting more of the expected emotion for that task) were viewed by parents as behaving in ways that were more similar to their typical behavior. The magnitudes of these effects were relatively modest, however.

Parent-Reported Typicality and Parent-Reported Child Emotions

We hypothesized that relatively lower parent-typicality ratings for lab tasks that elicited more negative emotions than for those designed to elicit positive emotions could be attributed to the somewhat more contrived nature of these tasks. As such, parents might base their typicality ratings less on a direct comparison of the observed behavior to prior examples occurring in highly similar contexts (as for positive tasks) and more on educated guesses based on their memory for the child's reactions to the most similar contexts she or he encountered in the past. If the parent's expectations for the child's reaction the task were violated (e.g., the child's emotional reaction differed from expectations), this should lead to lower ratings of typicality. To test this hypothesis, for each lab task, we compared children of parents who rated their child as exhibiting moderate to less typical behavior in that task (i.e., typicality rating ≤ 3) to children of parents who rated their child as exhibiting highly to extremely typical behavior in that task (i.e., typicality rating ≥ 4) on parent reports of the child's emotions in response to that task. This allowed us to test whether parent perceptions of the child's emotional reactions were driving their views of the typicality of the child's behavior.

For each task, ANOVAs were conducted with parent-reported child emotion variables as the dependent variables and low/high parent-reported typicality groups as the independent variable. Results are reported in Table 3. For the Box Empty task, which was designed to elicit sadness and anger, parents who rated their child as exhibiting moderate to less typical behavior rated their child as experiencing significantly less sadness than did parents who rated their child as exhibiting highly typical behavior. $F(1, 201) = 6.77, MSE = 42.3, p = .010$. Similar patterns were observed for parent reports of child anger, $F(1, 200) = 5.75, MSE = 40.2, p = .017$, and nervousness, $F(1, 199) = 4.61, MSE = 14.2, p = .033$. These findings suggest that children who responded less negatively to the

Table 3
Mean Differences in Parent-Reported Child Emotions Between Low and High Parent-Reported Typicality Groups

Variable	BX		DT		GC		TB		EO		SA	
	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High
Anger	5.00 (4.59)	7.39 (6.89)	3.10 (4.34)	3.59 (4.83)	5.01 (5.02)	4.83 (4.97)	5.46 (4.54)	6.81 (5.69)	1.14 (0.64)	1.23 (1.11)	2.17 (3.68)	2.06 (3.29)
Fearfulness	1.74 (2.36)	1.86 (2.65)	1.25 (0.87)	1.12 (0.62)	1.25 (1.15)	1.04 (0.31)	1.21 (0.91)	1.08 (0.52)	8.00 (6.97)	4.68 (4.50)	7.68 (7.07)	7.35 (6.62)
Sadness	7.70 (5.36)	10.4 (6.88)	4.69 (5.39)	4.95 (4.86)	2.29 (2.81)	2.80 (3.13)	3.98 (4.52)	4.43 (4.58)	1.48 (1.42)	1.29 (1.45)	3.89 (5.01)	4.21 (5.51)
Happiness	8.25 (4.38)	8.68 (5.31)	11.1 (6.23)	13.8 (6.57)	9.59 (5.88)	10.8 (6.27)	8.30 (4.70)	9.74 (5.81)	12.2 (6.44)	15.5 (6.36)	4.12 (3.42)	6.10 (5.29)
Surprise	8.11 (5.35)	9.18 (6.02)	4.68 (3.78)	4.99 (4.37)	2.49 (2.74)	3.60 (4.78)	3.81 (3.94)	4.31 (4.20)	7.00 (4.74)	7.53 (5.06)	6.29 (4.40)	6.36 (4.88)
Contentment	7.64 (4.91)	8.01 (5.41)	11.2 (6.84)	12.9 (6.48)	9.88 (5.97)	11.1 (6.93)	8.33 (5.69)	9.20 (5.51)	11.3 (6.80)	15.8 (6.67)	5.90 (4.56)	7.40 (5.83)
Nervousness	1.98 (2.19)	3.26 (4.22)	2.44 (3.46)	1.56 (1.52)	3.36 (4.16)	2.18 (2.53)	2.32 (3.00)	1.81 (1.97)	8.29 (6.62)	5.09 (5.30)	10.4 (7.40)	11.2 (7.87)

Note. Standard deviations are listed in parentheses below reported means. BX = Box Empty; DT = Disappointing Toy; GC = Green Circle; TB = Transparent Box; EO = Exploring New Objects; SA = Stranger Approach; Low = parent typicality ratings ≤ 3 ; High = parent typicality ratings ≥ 4 .

Box Empty task were viewed by their parents as behaving less typically than those children whose parents reported their behavior as more sad, angry, or nervous, consistent with an interpretation that the failure to mount an expected negative response was viewed as an aberration from typical behavior.

On other tasks designed to elicit sadness and anger, lower typicality ratings were associated with parents reporting their child as having a more negative/less positive response to the task. In the Disappointing Toy task, low and high parent-reported typicality groups differed on parent-reported child nervousness, $F(1, 194) = 6.15$, $MSE = 6.15$, $p = .014$, and happiness, $F(1, 200) = 7.41$, $MSE = 41.8$, $p = .007$. Parents who rated their child as exhibiting moderate to less typical behavior rated their child as displaying more nervousness and less happiness than their counterparts who were viewed as behaving more typically. A similar pattern was evident for the Green Circles task. Parents who rated their child as exhibiting moderate to less typical behavior in Green Circles rated their child as experiencing significantly more nervousness than did parents who rated their child as exhibiting highly typical behavior, $F(1, 195) = 6.14$, $MSE = 10.8$, $p = .014$. Thus, parent perceptions of child anxiety in response to these tasks seemed to drive parents' view that the child's behavior was not highly typical. Finally, for the Transparent Box task, parents who rated their child as exhibiting moderate to less typical behavior did not differ significantly in their report of their child's emotions from parents who rated their child as exhibiting highly typical behavior.

On tasks specifically designed to elicit nervousness and fear, lower typicality seemed to be related to parent perceptions of both lower positive emotions and higher fear. For Exploring New Objects, groups differed for happiness, $F(1, 198) = 9.29$, $MSE = 40.7$, $p = .003$; contentment, $F(1, 199) = 15.5$, $MSE = 44.9$, $p < .001$; nervousness, $F(1, 200) = 11.8$, $MSE = 31.8$, $p = .001$; and fearfulness $F(1, 200) = 15.0$, $MSE = 26.9$, $p < .001$. Parents who rated their child as exhibiting moderate to less typical behavior rated their child as displaying less happiness, less contentment, more nervousness, and more fearfulness than children perceived to behave more typically. This is consistent with the interpretation that children who responded more negatively and less positively to this task were viewed by their parents as behaving less typically than those children whose parents reported they were happier and less afraid. Similarly, results from the Stranger Approach task indicated that parents who rated their child as exhibiting moderate to less typical behavior rated their child as experiencing significantly less happiness than did parents who rated their child as exhibiting highly typical behavior, $F(1, 186) = 6.83$, $MSE = 23.0$, $p = .010$. However, the groups did not differ on parent-reported fear or nervousness in response to this task. These findings suggest that parents whose children were less happy in response to the stranger were viewed as behaving atypically from their normal behavior.

Thus, although parents generally viewed their child's responses to tasks designed to elicit negative emotions as generating behavior that was at least moderately typical of their child's usual behavior, those parents who did report their child's behavior to be less typical differed from parents who saw their child's behavior as more typical on their perceptions of their child's emotions during the task. For Box Empty, this finding was driven by the children described by parents as behaving less typically also seeming to be less upset by the task. For other negative emotion-eliciting tasks, parent perceptions of anxious reactions were associated with lower typicality. Given that parents may be less likely to observe fear in naturalistic settings (as it occurs

less frequently than other negative emotions), its occurrence in the novel lab tasks may have been somewhat surprising to parents, thus producing lower typicality ratings.

Parent-Reported Typicality and Objective Coding of Child Emotions

We hypothesized that relatively lower parent-typicality ratings for lab tasks designed to elicit negative emotions than for those designed to elicit positive emotions could also be attributed to the lower base rate of negative emotions observed in naturalistic settings. To test this hypothesis, for each lab task, we compared children of parents who rated their child as exhibiting moderate to less typical behavior in that task (i.e., typicality rating ≤ 3) to children of parents who rated their child as exhibiting highly to extremely typical behavior in that task (i.e., typicality rating ≥ 4) on objective coding of child emotion in response to that task.

For each task, ANOVAs were conducted with objective coding of child emotion variables as the dependent variables and low/high parent-reported typicality groups as the independent variable. Results are reported in Table 4. For Box Empty, Disappointing Toy, and Transparent Box, the three tasks designed to elicit sadness and anger, children who were rated as exhibiting moderate to less typical behavior by their parents did not differ significantly in their objectively coded emotions (i.e., PE, fear, sadness, anger) from children who were rated as exhibiting highly typical behavior.

However, on tasks designed to elicit fear and nervousness, lower parent-typicality ratings were related to higher objective coding of fear. For example, in Exploring New Objects, groups differed for objective coding of fear, $F(1, 199) = 7.00$, $p = .009$, $\eta^2 = .022$. Children who were rated as exhibiting moderate to less typical behavior were coded as displaying more expressions indicative of fear than did children perceived to behave more typically. This is consistent with the interpretation that children who responded more negatively to this task were viewed by their parents as behaving less typically than those children who had a less fearful response. In the second task designed to elicit fear, Stranger Approach, children who were rated as exhibiting moderate to less typical behavior by their parents did not differ significantly in their objectively coded emotions from children who were rated as exhibiting highly typical behavior.²

Discussion

We investigated the ecological validity of laboratory assessments of child temperament by examining the degree to which parents reported their child's responses to the lab tasks as typical of his or her usual behavior. The data in the present study were collected with the intention of capitalizing on the advantages of both parent-reported and laboratory measures of child temperament. The parents' extensive knowledge of their child's developmental and emotional history provided a unique approach to test-

² We also tested whether parent typicality moderated associations between parent reports of child emotion and objective coding of that emotion; there was no evidence that convergence was moderated by parent perceived typicality for PE/happiness, sadness, or anger. It did moderate convergence for fear such that parent-coder agreement for fear was higher for children whose parents reported their behavior was more typical, but the effect size of this interaction was modest.

Table 4
Mean Differences in Objective Coding of Child Emotions Between Low and High Parent-Reported Typicality Groups

Variable	BX		DT		GC		TB		EO		SA	
	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High
Anger	8.13 (11.5)	11.7 (13.8)	2.60 (4.45)	3.64 (6.87)	4.54 (6.61)	5.85 (9.13)	7.01 (8.27)	6.80 (8.36)	0.70 (1.11)	0.76 (2.27)	0.91 (2.71)	0.65 (1.62)
Fearfulness	0.32 (1.44)	0.46 (1.74)	0.03 (0.25)	0.21 (0.84)	0.12 (0.46)	0.20 (0.83)	0.09 (0.52)	0.04 (0.32)	27.4 (28.5)	15.9 (25.2)	4.73 (4.71)	5.21 (6.21)
Sadness	13.4 (13.6)	14.0 (17.0)	4.88 (7.48)	6.58 (10.1)	2.63 (3.72)	3.24 (5.64)	5.69 (6.48)	7.28 (10.1)	1.40 (4.86)	1.33 (7.89)	1.96 (3.87)	2.73 (5.87)
Happiness	25.4 (20.9)	25.8 (17.3)	48.4 (31.6)	55.6 (33.4)	28.4 (21.4)	28.0 (21.1)	34.4 (30.9)	29.0 (23.0)	40.3 (34.2)	44.7 (34.7)	10.5 (11.3)	11.2 (14.3)
Surprise	0.97 (2.09)	0.97 (1.96)	0.20 (0.51)	0.31 (0.91)	0.04 (0.26)	0.32 (1.00)	0.11 (0.44)	0.24 (0.82)	0.40 (1.32)	0.84 (1.88)	0.21 (0.73)	0.05 (0.30)

Note. Standard deviations are listed in parentheses below reported means. BX = Box Empty; DT = Disappointing Toy; GC = Green Circle; TB = Transparent Box; EO = Exploring New Objects; SA = Stranger Approach; Low = parent typicality ratings ≤ 3; High = parent typicality ratings ≥ 4.

ing the ecological validity of laboratory assessments of temperament. Parents were asked to report directly on the degree to which their child’s behavior during each of the 10 lab tasks varying in their potency for eliciting different temperament traits was representative of his or her usual behavior. This allowed us to test the common assumption that the contrived scenarios and novel settings in laboratory assessments preclude children from displaying emotional behavior reflective of their temperament as exhibited in their natural environment. Results provided evidence to challenge this assumption, indicating that, for the most part, parents rated their child’s behavior as highly typical of his or her usual behavior.

Consistent with our hypotheses, we found that parents viewed their child’s behavior as more typical of his or her usual behavior during tasks that elicited higher levels of PE than during tasks that elicited higher levels of NE. This finding was consistent with previous research suggesting that parents rate their child’s positive attributes as more consistently true than they do negative ones (e.g., Gretarsson & Gelfand, 1988). Moreover, according to Funder’s (1995) realistic accuracy model of accurate personality judgment, “good targets” of judgment (i.e., people who are judged most accurately) are those whose traits are made more available by their saliency. This is frequently true for individuals with high positive mood, sociability, and greater activity level. Thus, when children exhibit greater happiness, they are perceived by parents as engaging in more typical behavior. It is also possible that our finding is a result of the lower base rate of negative (compared to positive) emotions in naturalistic contexts; parents may observe negative emotions less frequently in their child and may therefore be more likely to attribute the occurrence of negative emotions to situational characteristics rather than to stable dispositions of the child.

The mechanisms underlying these findings must be examined in future research. It is possible that parents observe certain negative emotions such as fear less frequently because they are aware of situations of which their child is fearful and therefore avoid these contexts, leaving fewer opportunities to observe variations in child fear. Social desirability may also influence their ratings of typicality of the behavior observed in laboratory settings. Parents may have a desire for their child to be on his or her best behavior for the lab visit and may therefore rate their child’s expressions of PE as more typical, because PE responses are presumably more socially desirable than NE responses. There is an extensive literature on response bias and social desirability (e.g., Edwards, 1957), but to our knowledge no studies have explored parental social desirability and its impact on parent ratings of child behavior. Other researchers (De Los Reyes & Kazdin, 2005) have proposed theoretical models such as the attrition bias context (ABC) model to provide a framework for understanding informant discrepancies in assessment of child psychopathology. The ABC framework could also be used to conceptualize and test mechanisms driving the present findings, such as considering both parental characteristics (e.g., personality, psychopathology, acceptance, stress) and perception (e.g., attributions, memory recall), and their impact on parent-reported typicality.

Bivariate correlation analyses revealed that objective markers of more surgent and extraverted behavior (i.e., PE, anticipatory PE, interest, initiative, and sociability) were associated with higher parent ratings of typicality. However, objective markers of NE

were not associated with the extent to which parents rated their child as exhibiting typical behavior during the lab tasks. Consistent with these results, within-subject correlations showed that parents viewed their child as exhibiting more typical behavior during tasks in which she or he exhibited higher levels of PE. Surprisingly, results also indicated that parents viewed their child as exhibiting more typical behavior in tasks in which she or he exhibited higher levels of fear. Additional analyses indicated that this association was observed only in tasks that were designed to elicit fearful emotions. Consistent with their design, the highest mean levels of coded fear occurred in tasks that involved exposing children to unfamiliar adults and stimuli. These tasks presumably were viewed by parents as potentially anxiety-provoking scenarios. Overall, children who were more responsive to lab tasks (i.e., who exhibited greater frequency and intensity of emotions, showed more engagement and interest, and were more active) were viewed by their parents as exhibiting more typical behaviors. These associations tended to be stronger in tasks designed to elicit positive emotions, with the exception of fear, where parents viewed their child as behaving more typically when the child exhibited higher levels of fear during tasks that were designed to elicit nervousness or fear.

Our findings also suggested that lower parent-typicality ratings across lab tasks designed to elicit more negative emotions could be attributed to parents expecting their child to have reacted differently in a more naturalistic setting. More specifically, results indicated that this could manifest in two ways. For example, on the Box Empty task, parents who viewed their child as exhibiting moderate to less typical behavior rated their child as experiencing significantly less sadness, anger, and nervousness than did parents who viewed their child as exhibiting highly typical behavior. These results supported the hypothesis that parents who viewed their child as displaying less typical behavior on this task may have expected their child to have a more negative emotional reaction. A different pattern was observed for other negative tasks (Disappointing Toy, Green Circles), wherein parents who viewed their child as exhibiting moderate to less typical behavior viewed their child as experiencing more nervousness than did parents who viewed their child as exhibiting highly typical behavior. This may be because parents expected their child to respond to these tasks with sadness or anger but with not fear. This was consistent with the design of these tasks, which were meant primarily to elicit sadness and anger. Two tasks were specifically intended to elicit fear/nervousness: Stranger Approach and Exploring New Objects. Parent typicality was not associated with parent reports of child fear/nervousness in the former, but it was in the latter. Children rated by their parents as more fearful during Exploring New Objects were also seen by parents as exhibiting less typical behavior. This suggested that parents might have been surprised by their child's fearful reaction to novel objects but were not surprised by such a reaction to an encounter with a strange adult.

These findings also suggested that lower parent-typicality ratings across lab tasks eliciting more negative emotions could potentially be attributed to a violation of the parents' expectations of their child's typical behavior in either direction (i.e., expecting their child to have a less negative and/or more positive reaction, or vice versa). In general, parents who rated their child as behaving less typically also viewed their child as exhibiting more nervousness rather than sadness or anger. One possible explanation is that

fear occurs at a lower base rate. In other words, the presence of nervous behavior may have been more surprising to parents and prompted lower typicality ratings, whereas sadness and anger are more common emotions, such that parents find their child's sadness or anger in response to a task meant to elicit that reaction as more typical. When we examined the objective coding of children's emotions to tasks designed to elicit negative emotions, children who were rated as exhibiting moderate to less typical behavior were coded as displaying more expressions indicative of fear than were children perceived to behave more typically. It is possible that parents may not have as much experience observing their child in similar situations as those in the fear-eliciting tasks and thus may have been surprised by the degree of fear the child exhibited. However, this finding was also specific to one task designed to elicit fear. On the remaining tasks designed to elicit negative emotions, children who were rated as behaving less typically and children who were rated as exhibiting highly typical behavior did not differ significantly in their objectively coded emotions.

The present study has a number of strengths, as it capitalized on the advantages of both laboratory assessments and parent questionnaire methods. Other advantages of this approach included asking parents to rate how typical their child's behavior was during the lab task immediately following the termination of each lab task, which may have helped reduce memory biases associated with parent reports. It is possible that parent report and lab methods of child traits diverge not because the behavior generated in response to lab tasks elicited is unusual or unrepresentative of typical child behavior but rather because parent reports and other measurement approaches (e.g., objective coders) interpret this behavioral evidence differently when using it to indicate a child's relative standing on a trait.

The study also has some notable limitations. Our findings speak only to parent perceptions of the ecological validity of the particular tasks used in our battery. Different tasks or approaches to assessing child traits (e.g., using a single task or shorter battery) may not be viewed by parents as eliciting behavior that is highly typical. Because responses to any task are never determined solely by the action of a single temperament trait, some tasks can provide more or less accurate estimates of particular traits. For example, children high in behavioral inhibition show reduced approach behavior in laboratory contexts involving the presentation of novel stimuli (those designed to elicit fear or anxiety), but they do not differ from children low in this trait on their level of approach in non-novel laboratory tasks (Laptook et al., 2008; Laptook, Klein, Olino, Dyson, & Carlson, 2010). Additionally, parent perception of whether their child's behavior during the laboratory task was typical of his or her usual behavior was assessed with only one item, so typicality ratings for individual tasks have unknown reliability. However, the internal consistency of the parent-typicality ratings across all 10 tasks was found to be quite high ($\alpha = .82$). Nonetheless, future studies should explore collecting multiple items to assess typicality of child behavior. Our findings also may not generalize to children younger or older than the age group in our sample. Therefore, future studies should investigate the ecological validity of laboratory methods in other developmental periods such as infancy, young toddlerhood, middle childhood, and adolescence.

Despite these limitations, our results support the use of laboratory measures as complementary to parent report measures and as ecologically valid assessments of child temperament. These findings also raise a number of important questions about the assessment of child temperament and ecological validity of laboratory methods. For example, given the extensive literature suggesting that parental characteristics influence their report of child behavior (e.g., De Los Reyes & Kazdin, 2005; Durbin & Wilson, 2012; Foley et al., 2005; Gartstein et al., 2009), future studies should investigate other biases and the underlying mechanisms that may lessen or improve the accuracy of parent report. Future research should also examine which specific factors contribute to and predict higher typicality ratings, both to bolster understanding of the features of a child's responses across different lab tasks that contribute to parents' interpretation and understanding of their own child's behavior and to refine laboratory approaches to maximize the fidelity with which they capture children's temperament traits.

References

- Bates, J. E. (1994). Parents as observers of their children's development. In S. Friedman & H. C. Haywood (Eds.), *Developmental follow-up: Concepts, domains, and methods* (pp. 197–216). San Diego, CA: Academic Press.
- Boyle, M. H., & Pickles, A. (1997). Maternal depressive symptoms and ratings of emotional disorder symptoms in children and adolescents. *Journal of Child Psychology and Psychiatry*, *38*, 981–992. <http://dx.doi.org/10.1111/j.1469-7610.1997.tb01615.x>
- Caspi, A., Moffitt, T. E., Newman, D. L., & Silva, P. A. (1996). Behavioral observations at age 3 years predict adult psychiatric disorders. *Archives of General Psychiatry*, *53*, 1033–1039. <http://dx.doi.org/10.1001/archpsyc.1996.01830110071009>
- Caspi, A., & Roberts, B. W. (2005). Personality development: Stability and change. *Annual Review of Psychology*, *56*, 453–484. <http://dx.doi.org/10.1146/annurev.psych.55.090902.141913>
- Chi, T. C., & Hinshaw, S. P. (2002). Mother-child relationships of children with ADHD: The role of maternal depressive symptoms and depression-related distortions. *Journal of Abnormal Child Psychology*, *30*, 387–400. <http://dx.doi.org/10.1023/A:1015770025043>
- Christensen, A., Margolin, G., & Sullaway, M. (1992). Interparental agreement on child behavior problems. *Psychological Assessment*, *4*, 419–425. <http://dx.doi.org/10.1037/1040-3590.4.4.419>
- De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, *131*, 483–509. <http://dx.doi.org/10.1037/0033-2909.131.4.483>
- Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin*, *130*, 959–988. <http://dx.doi.org/10.1037/0033-2909.130.6.959>
- Dougherty, L. R., Klein, D. N., Durbin, C. E., Hayden, E. P., & Olino, T. M. (2010). Temperamental positive and negative emotionality and children's depressive symptoms: A longitudinal prospective study from age three to age ten. *Journal of Social and Clinical Psychology*, *29*, 462–488. <http://dx.doi.org/10.1521/jscp.2010.29.4.462>
- Duhig, A. M., Renk, K., Epstein, M. K., & Phares, V. (2000). Interparental agreement on internalizing, externalizing, and total behavior problems: A meta-analysis. *Clinical Psychology: Science and Practice*, *7*, 435–453. <http://dx.doi.org/10.1093/clipsy.7.4.435>
- Dunn, L. M., & Dunn, L. M. (1997). *Examiner's manual for the PPVT-III: Peabody Picture Vocabulary Test, Third Edition*. Circle Pines, MN: American Guidance Services.
- Durbin, C. E. (2010). Modeling temperamental risk for internalizing psychopathology using developmentally sensitive laboratory paradigms. *Child Development Perspectives*, *4*, 168–173. <http://dx.doi.org/10.1111/j.1750-8606.2010.00137.x>
- Durbin, C. E., Hayden, E. P., Klein, D. N., & Olino, T. M. (2007). Stability of laboratory-assessed temperamental emotionality traits from ages 3 to 7. *Emotion*, *7*, 388–399. <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. <http://dx.doi.org/10.1037/0021-843X.114.1.28>
- Durbin, C. E., & Wilson, S. (2012). Convergent validity of and bias in maternal reports of child emotion. *Psychological Assessment*, *24*, 647–660. <http://dx.doi.org/10.1037/a0026607>
- Dyson, M. W., Olino, T. M., Durbin, C. E., Goldsmith, H. H., & Klein, D. N. (2012). The structure of temperament in preschoolers: A two-stage factor analytic approach. *Emotion*, *12*, 44–57. <http://dx.doi.org/10.1037/a0025023>
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. New York, NY: Dryden.
- Foley, D. L., Rutter, M., Angold, A., & Pickles, A. (2005). Making sense of informant disagreement for overanxious disorder. *Journal of Anxiety Disorders*, *19*, 193–210. <http://dx.doi.org/10.1016/j.janxdis.2004.01.006>
- Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, *102*, 652–670. <http://dx.doi.org/10.1037/0033-295X.102.4.652>
- Gartstein, M. A., Bridgett, D. J., Dishion, T. J., & Kaufman, N. K. (2009). Depressed mood and maternal report of child behavior problems: Another look at the depression-distortion hypothesis. *Journal of Applied Developmental Psychology*, *30*, 149–160. <http://dx.doi.org/10.1016/j.appdev.2008.12.001>
- Goldsmith, H. H., Buss, A. H., Plomin, R., Rothbart, M. K., Thomas, A., Chess, S., . . . McCall, R. B. (1987). Roundtable: What is temperament? Four approaches. *Child Development*, *58*, 505–529. <http://dx.doi.org/10.2307/1130527>
- Goldsmith, H. H., Reilly, J., Lemery, K. S., Longley, S., & Prescott, A. (1995). *Laboratory Temperament Assessment Battery: Preschool version*. Unpublished manuscript.
- Gretarsson, S. J., & Gelfand, D. M. (1988). Mothers' attributions regarding their children's social behavior and personality characteristics. *Developmental Psychology*, *24*, 264–269. <http://dx.doi.org/10.1037/0012-1649.24.2.264>
- Hayden, E. P., Klein, D. N., Durbin, C. E., & Olino, T. M. (2006). Positive emotionality at age 3 predicts cognitive styles in 7-year-old children. *Development and Psychopathology*, *18*, 409–423. <http://dx.doi.org/10.1017/S0954579406060226>
- Kagan, J. (1998). Biology and the child. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (5th ed., pp. 177–235). New York, NY: Wiley.
- 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. <http://dx.doi.org/10.1017/S0954579402003048>
- Kochanska, G., Aksan, N., & Carlson, J. J. (2005). Temperament, relationships, and young children's receptive cooperation with their parents. *Developmental Psychology*, *41*, 648–660. <http://dx.doi.org/10.1037/0012-1649.41.4.648>
- Kochanska, G., Murray, K., Jacques, T. Y., Koenig, A. L., & Vandegest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development*, *67*, 490–507. <http://dx.doi.org/10.2307/1131828>
- Laptook, R. S., Klein, D. N., Durbin, C. E., Hayden, E. P., Olino, T. M., & Carlson, G. (2008). Differentiation between low positive affectivity and behavioral inhibition in preschool-age children: A comparison of

- behavioral approach in novel and non-novel contexts. *Personality and Individual Differences*, 44, 758–767. <http://dx.doi.org/10.1016/j.paid.2007.10.010>
- Laptook, R. S., Klein, D. N., Olino, T. M., Dyson, M. W., & Carlson, G. (2010). Low positive affectivity and behavioral inhibition in preschool-age children: A replication and extension of previous findings. *Personality and Individual Differences*, 48, 547–551. <http://dx.doi.org/10.1016/j.paid.2009.12.003>
- Lemery, K. S., Goldsmith, H. H., Klinnert, M. D., & Mrazek, D. A. (1999). Developmental models of infant and child temperament. *Developmental Psychology*, 35, 189–204. <http://dx.doi.org/10.1037/0012-1649.35.1.189>
- Majdandžić, M., & Van Den Boom, D. C. (2007). Multimethod longitudinal assessment of temperament in early childhood. *Journal of Personality*, 75, 121–168. <http://dx.doi.org/10.1111/j.1467-6494.2006.00435.x>
- Mesman, J., & Koot, H. M. (2000). Child-reported depression and anxiety in preadolescence: II. Preschool predictors. *Journal of the American Academy of Child & Adolescent Psychiatry*, 39, 1379–1386. <http://dx.doi.org/10.1097/00004583-200011000-00012>
- Mitchell, S. K. (1979). Interobserver agreement, reliability, and generalizability of data collected in observational studies. *Psychological Bulletin*, 86, 376–390. <http://dx.doi.org/10.1037/0033-2909.86.2.376>
- Olino, T. M., Klein, D. N., Dyson, M. W., Rose, S. A., & Durbin, C. E. (2010). Temperamental emotionality in preschool-aged children and depressive disorders in parents: Associations in a large community sample. *Journal of Abnormal Psychology*, 119, 468–478. <http://dx.doi.org/10.1037/a0020112>
- Patrick, C. J., Curtin, J. J., & Tellegen, A. (2002). Development and validation of a brief form of the Multidimensional Personality Questionnaire. *Psychological Assessment*, 14, 150–163.
- Pedlow, R., Sanson, A., Prior, M., & Oberklaid, F. (1993). Stability of maternally reported temperament from infancy to 8 years. *Developmental Psychology*, 29, 998–1007. <http://dx.doi.org/10.1037/0012-1649.29.6.998>
- Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, 126, 3–25. <http://dx.doi.org/10.1037/0033-2909.126.1.3>
- 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. <http://dx.doi.org/10.1111/1467-8624.00355>
- Rothbart, M. K., & Goldsmith, H. H. (1985). Three approaches to the study of infant temperament. *Developmental Review*, 5, 237–260. [http://dx.doi.org/10.1016/0273-2297\(85\)90012-7](http://dx.doi.org/10.1016/0273-2297(85)90012-7)
- 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. <http://dx.doi.org/10.2307/1131512>
- Shiner, R. L. (2000). Linking childhood personality with adaptation: Evidence for continuity and change across time into late adolescence. *Journal of Personality and Social Psychology*, 78, 310–325. <http://dx.doi.org/10.1037/0022-3514.78.2.310>
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428. <http://dx.doi.org/10.1037/0033-2909.86.2.420>
- Webster-Stratton, C., & Hammond, M. (1988). Maternal depression and its relationship to life stress, perceptions of child behavior problems, parenting behaviors, and child conduct problems. *Journal of Abnormal Child Psychology*, 16, 299–315. <http://dx.doi.org/10.1007/BF00913802>
- 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. <http://dx.doi.org/10.1111/j.1467-6494.2012.00789.x>
- Youngstrom, E., Izard, C., & Ackerman, B. (1999). Dysphoria-related bias in maternal ratings of children. *Journal of Consulting and Clinical Psychology*, 67, 905–916. <http://dx.doi.org/10.1037/0022-006X.67.6.905>
- Youngstrom, E., Loeber, R., & Stouthamer-Loeber, M. (2000). Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. *Journal of Consulting and Clinical Psychology*, 68, 1038–1050. <http://dx.doi.org/10.1037/0022-006X.68.6.1038>

Received November 1, 2013

Revision received August 21, 2014

Accepted September 3, 2014 ■