

The Reliability and Validity of the Infant Behavior Questionnaire-Revised

By: Stephanie H. Parade and Esther M. Leerkes

Parade, S.H., & Leerkes, E.M. (2008). The reliability and validity of the infant behavior questionnaire-revised. *Infant Behavior and Development*, 31(4), 637-646.
DOI:10.1016/j.infbeh.2008.07.009.

*****Note: This version of the document is not the copy of record. Made available courtesy of Elsevier. Link to DOI: <http://dx.doi.org/10.1016/j.infbeh.2008.07.009>**

Abstract:

The reliability and validity of the Infant Behavior Questionnaire-Revised was examined in a sample of 6-month-old infants and their parents. One hundred and fifteen mothers and 79 fathers completed the IBQ-R and a measure of depression and 98 infants participated in a laboratory assessment of temperament. Internal consistency reliability was adequate for all 14 IBQ-R subscales for both mothers and fathers and inter-rater reliability of mother and father reports was demonstrated for 11 of 14 subscales. Convergent validity was established between observed fear and mother reported fear and father reported approach. Parent depression and infant gender were examined as moderators of the concordance between parent reported and observed temperament. As predicted, concordance was higher when parents reported low versus high symptoms of depression. Infant gender did not alter concordance.

Article:

Since Thomas, Chess, and Birch's (1968) seminal work on infant temperament, interest in the measurement of temperament has increased steadily. Evidence that infant temperament predicts parental well-being and parenting behavior (Crockenberg & Leerkes, 2003; Paulussen-Hoogeboom, Stams, Hermanns, & Peetsma, 2007) as well as subsequent child adjustment (Calkins & Degnan, 2006; Rothbart & Bates, 1998) underscores the importance of measuring temperament effectively. Infant temperament is frequently studied using parent reports, yet the validity and reliability of parent reports has been a topic of much debate in the infancy literature (Rothbart & Bates). Many have raised concerns that parents' reports can be biased (Forman et al., 2003; Vaughn, Bradley, Joffe, Seifer, & Barglow, 1987), whereas others have argued that parents' reports are comprised of objective components (child's actual behaviors) more than subjective components (parents' perceptions and biases; Bates, 1980). In this study, we examine the reliability and validity of the Infant Behavior Questionnaire-Revised (IBQ-R; Gartstein & Rothbart, 2003) in a sample of 6-month-old infants and their parents. Specifically, we investigate internal consistency reliability for mother and father reports, inter-rater reliability between mother and father reports, and convergent validity of the IBQ-R with observed indices of temperament. Finally, we investigate the possibility that parental depression and infant gender moderate the convergence of parent reports and observed indices of temperament.

The Infant Behavior Questionnaire (IBQ) was originally developed in 1981 by Mary Rothbart as a parent-report measure of infant temperament. Recently, Gartstein and Rothbart (2003) revised

the measure to reflect findings from contemporary research pertaining to infant temperament. Rather than capturing a broad infant trait such as difficulty, the IBQ-R measures specific dimensions of temperament. The IBQ-R is composed of 14 subscales including approach, vocal reactivity, high pleasure, smile and laughter, activity level, perceptual sensitivity, sadness, distress to limitations, fear, falling reactivity, low pleasure, cuddliness, duration of orienting, and soothability. Parents are asked to report on specific infant behaviors during specific events in the last two weeks (e.g. when introduced to an unfamiliar adult, how often did the baby cling to a parent), using a 7-point Likert scale with response options that range from never (1) to always (7), as well as does not apply if the event did not occur within the time span of interest. The IBQ-R differs from the original IBQ in that it includes eight new subscales: approach, vocal reactivity, high pleasure, perceptual sensitivity, sadness, falling reactivity, low pleasure, and cuddliness. Additionally, the original IBQ subscales were refined in light of recent research on temperament. For example, advances made in the measurement of temperament in childhood (Rothbart, Ahadi, Hershey, & Fisher, 2001) informed the revision of the original IBQ.

RELIABILITY OF THE IBQ-R

Internal reliability and inter-rater reliability of the IBQ-R have been previously investigated by Gartstein and Rothbart (2003). Although internal reliability was at acceptable levels for all 14 IBQ-R subscales ranging from .70 to .90 for parents whose children were between the ages of 3 and 9 months, these values were not reported separately for mothers and fathers. Thus, it is important to demonstrate the scale has similar properties for both mothers and fathers given each are frequently asked to rate their infant's temperament. Because of the small sample of secondary caregivers in Gartstein and Rothbart's study ($n = 26$), inter-rater reliability for many of the IBQ-R subscales was relatively low with only six subscales having significant correlations between mother and father reports. Further, their sample of secondary caregivers included both fathers (58%) and other unspecified secondary caregivers. Assuming other secondary caregivers observe children in different contexts than parents (e.g., in childcare settings with multiple adults and children), we anticipate that inter-rater reliability will be higher when both respondents are parents.

VALIDITY OF THE IBQ-R

Although convergent validity of the IBQ-R and observed indices of temperament has not yet been examined, Gartstein and Rothbart (2003) investigated discriminant validity of the IBQ-R by considering correlations between IBQ-R subscales and found some evidence for the independence of each. Convergent validity of prior versions of the IBQ (Rothbart, 1981) with observed indices of temperament has been demonstrated; however associations have been relatively modest (Forman et al., 2003). Further, parent and other child characteristics have been found to alter the concordance between those variables as discussed below (Forman et al., 2003; Gill & Link, 2000; Leerkes & Crockenberg, 2003).

PARENT REPORTS OF TEMPERAMENT AND PARENT CHARACTERISTICS

Parent reports of infant temperament may be influenced by parent characteristics, particularly in high risk samples. Prior research has demonstrated that in normative samples, observer and father ratings of temperament explain more variance in maternal reports of temperament than maternal characteristics (Bates & Bayles, 1984), suggesting maternal reports are a valid measure of temperament. However, in high-risk samples, maternal characteristics have been found to be

more predictive of maternal reports of temperament than observed infant behavior (Sameroff, Seifer, & Elias, 1982) suggesting that some mothers may be less accurate in their reports of temperament than other mothers. Consistent with this view, concordance between maternal reports and observed indices of temperament have been found to be higher when mothers report less stress and hostility and when they have daughters (Gill & Link, 2000), when they report having their emotional needs met in childhood (Leerkes & Crockenberg, 2003), and when they exhibit low depressive symptoms (Forman et al., 2003; Leerkes & Crockenberg). In this study, we expand upon prior research by examining moderating effects of parent depression and infant gender on the concordance between both maternal and *paternal* reports of temperament and observed indices of temperament.

Previous research has demonstrated that depressed mothers have more difficulty distinguishing between their infants' cries and tend to rate their infants as more difficult than less depressed women (Donovan, Leavitt, & Walsh, 1998; Mebert, 1991; Schuetze & Zeskind, 2001). Similar studies have found that fathers who were more depressed tended to rate their infants as fussier and more difficult (Atella, DiPietro, Smith, & St. James-Roberts, 2003; Dave, Nazareth, Sherr, & Senior, 2005). These findings are consistent with the view that depressed parents may misinterpret infant signals because of a preoccupation with their own negative feelings and because of the pattern of attributions that characterize depression. Depressed parents may also ignore infant distress if it arouses feelings of anxiety and hopelessness. Both processes may undermine the accuracy of parents' reports of temperament.

Gender biases held by parents may also influence their reports of infant temperament. Mothers have been found to rate their sons higher on frustration than daughters even when there is no difference in observed measures of emotionality (Diener & Bradshaw, 2002; Polak, Henderson, & Fox, 2002), a finding which is consistent with the view that children's gender influences parents' perceptions of temperament. Specifically, parents may over-rate anger in their sons and under-rate anger in their daughters given our society's view that men are more likely to express anger than women (Plant, Hyde, Keitner, & Devine, 2000). Providing support for this view, Gill and Link (2000) found that concordance between maternal reports and observed indices of frustration was higher for girls than boys. Likewise, parents may underrate fear in boys given evidence that it is less socially acceptable for boys to express fear than girls (Brody & Carter, 1982). Finally, given evidence that fathers are more likely to engage their children in a gender-typed manner ([Antill, 1987] and [Siegal, 1987]), and that boys who stray from gender norms are more negatively regarded than girls who stray from gender norms (Sandnabba & Ahlberg, 1999), it may be that fathers are more prone to a gender bias in rating temperament than are mothers.

Concordance between parent reports and observed indices of temperament appear to be strongest when well-established parent report measures are used, mothers and observers rate infant behavior in similar situations, and comparisons are made between the same dimensions of temperament (Rothbart & Bates, 1998). We address these issues by using the IBQ-R, a well-established temperament measure and observing temperament in laboratory situations that are conceptually similar to specific subscales of the IBQ-R. That is, exposure to a novel toy overlaps with the fear and approach subscale and an arm restraint procedure that limits infant movement and engagement overlaps with the distress to limitations subscale. Finally, we compare related

dimensions of infant behavior; reported fear and approach and observed distress to novelty; reported and observed distress to limitations. In sum, we hypothesize that:

1. Internal reliability for each IBQ-R subscale for both mothers and fathers will be acceptable.
2. Mother and father reports of infant temperament will correlate positively. We will explore the possibility that infant gender will moderate the strength of the relationship between mother and father reports such that agreement will be lower among parents of male infants.
3. Reported distress to novelty will correlate positively with observed fear whereas reported approach will correlate negatively with observed fear. And reported and observed distress to limitations will correlate positively.
4. Depression and child gender will moderate associations between parent reports and observed temperament. Specifically, concordance between observed temperament and reported temperament will be weaker for parents who are depressed and for those parents whose infants are male.

METHOD

Participants

One hundred and fifteen mothers and 79 fathers completed the IBQ-R. Mothers' mean age was 28 (range, 15–38), 67% had college degrees, and 77% were White. Fathers' mean age was 31 (range, 21–43), 67% had college degrees, and 84% were White. The mean family income was \$70,000 (range, \$6000–\$190,000). All infants were full-term and healthy; 56% were male. Ninety-eight of these mothers and infants participated in a laboratory observation of infant temperament. Infants who did not complete the laboratory observation were more likely to have mothers who were white ($\chi^2 = 5.83, p < .05$), and had fathers who rated them higher on three dimensions of temperament including smiling and laughter ($t = -2.43, p < .05$), high pleasure ($t = -3.35, p < .01$) and approach ($t = -2.37, p < .05$). Thus, there is no evidence that infants who were perceived to be more reactive were less likely to participate in the observation of temperament. All data collected from mothers was complete; however, one father did not complete four of the 14 IBQ-R subscales.

Procedure

Mothers and fathers were recruited at birthing classes in local hospitals and the public health department as part of a larger study about the origins of maternal sensitivity. Interested parents were contacted by telephone and given more details about the study. Participants were mailed a demographic questionnaire and consent form during the prenatal period. An observation of infant temperament was scheduled within 1 week of the infant's 6-month birthday. Mothers and fathers were mailed questionnaires including a measure of infant temperament and parent depression to be completed prior to the visit. Mothers and fathers were instructed not to discuss their questionnaire responses with one another. Families received a \$20 gift certificate for completing the 6-month measures.

Measures

The Center for Epidemiologic Studies-Depression Scale (CES-D). Depressive symptoms were assessed using this 20-item checklist of moods, feelings, and cognitions associated with

depression (e.g., I felt depressed, I felt that people dislike me) designed for use with community samples (Radloff, 1977). Respondents indicate how often they felt a particular way during the previous week on a 4-point scale. The CES-D demonstrates convergent validity with the Research Diagnostic Criteria, a standardized psychiatric interview, and with the Beck Depression Inventory (Spitzer, Endicott, & Robins, 1978). Items from the CES-D were averaged to derive measures of depressive symptoms for use in data analyses. Cronbach's $\alpha = .85$ for mothers and $.82$ for fathers.

The Infant Behavior Questionnaire-Revised (IBQ-R; Gartstein & Rothbart, 2003). The IBQ-R is a 191 item measure, organized into 14 subscales (activity level, distress to limitations, fear, duration of orienting, smile/laughter, high pleasure, low pleasure, soothability, falling reactivity, cuddliness, perceptual sensitivity, sadness, approach, and vocal reactivity), that is designed to assess infant temperament between the ages of three and twelve months. Mothers and fathers rated the frequency of infant behaviors on a scale from a 1 (never) to 7 (always). Internal reliability has previously been demonstrated to be good for each of the subscales (Gartstein & Rothbart, 2003). Items from each subscale were averaged to obtain scores. Internal consistency for each scale for mothers and fathers is discussed below.

Temperament observations. At 6 months of age infant behavior was videotaped to assess infant temperament in a laboratory observation adapted from Goldsmith and Rothbart (1996). Mothers brought their infants into the laboratory and were instructed to engage in five minutes of free play to acclimate themselves and their infants to the room. During the free play session the experimenter left the mothers and infants alone in the room. After the experimenter returned to the room, mothers were instructed to place their infants in a car seat and then sit beside them in a chair visible to the infant with some effort. Two 4-min tasks were then administered to elicit a fear response (novelty task) and an anger response (limiting task). Half of the infants engaged in the novelty task first, the other half engaged in the limiting task first. Between each task there was a 5-min break for mothers to comfort their infants.

Novelty task. During the novelty task a large table with wooden barriers on each end was placed in front of the infant in the car seat. A toy dump truck with loud sounds (beeping horn, engine noises, the phrases “start your engine!” and “load up the truck!”) was then placed on the table. From underneath the table the experimenter utilized a remote control to turn the truck on, to move it forward to approach the infant and backward away from the infant, to play the noises, and to turn the truck off at the end of the task. The experimenter was not visible to the infant. For the first three minutes of the fear task the truck was turned on and in motion, for the last minute the truck was turned off but left in front of the infant within arms reach. For the first minute of the task, mothers were instructed to not interact with their infant, for the remaining three minutes mothers were instructed to interact with their child in any manner they wished other than touching the toy or removing their infant from the seat.

Limitations task. During the limitations task the experimenter knelt in front of the infant who was in the car seat and gently held down the infant's forearms so that they were immobile. During all four minutes of the task the experimenter kept her head down and did not interact with the infant. At the end of the four minutes the experimenter released the infant's forearms. During the first minute, mothers were instructed to remain neutral and uninvolved. Following the first minute,

mothers were instructed to interact in any manner they wished other than removing their infant from the seat or touching the experimenter.

Coding infant reactivity and discrete behaviors. Videotapes were subsequently coded by graduate and undergraduate students for infant affect and emotion-related behaviors. Infant affect was continuously rated on a 7-point scale from 1 (high positive affect) to 7 (high negative affect) based on the infant's facial expressions, body tension, and vocalizations adapted from Braungart-Rieker and Stifter (1996). Inter-rater reliability was calculated based on 33 videotapes that were double coded; kappa for infant reactivity was .73. The average level of affect during each task was calculated; high scores indicate greater distress. Discrete infant behavioral responses to each task were continuously coded in the categories of body position (neutral body position, approach stimulus, withdraw from stimulus, and resist stimulus), gaze (look at stimulus, look away from stimulus at other object or mother, and eyes closed), cry, and startle. Inter-rater reliability was calculated based on 22 videotapes that were double coded; kappa for discrete behavioral categories ranged from .72 to .93. The percent of time the infant engaged in each of these behaviors was calculated separately for the novelty task and limits task.

For data analysis, infant affect ratings and the percent of time the infant engaged in each of the behaviors that reflect heightened arousal were standardized and averaged to create composite scores for observed fear and observed anger. Observed fear was the composite of average affect, cry, eyes closed, look at stimulus (reverse scored), look away from stimulus at other object or mother, approach (reverse scored), withdraw, and startle during the novelty task ($\alpha = .62$). Observed anger was the composite of average affect, cry, resist, and eyes closed during the limits task ($\alpha = .70$).

RESULTS

Preliminary analyses

As demonstrated in Table 1, descriptive statistics were calculated for all variables. *t*-tests were conducted for each of the variables based on race. Minority mothers reported more depression ($M = 1.54$, $S.D. = .48$) than White mothers ($M = 1.37$, $S.D. = .30$; $t(113) = -2.18$, $p < .05$) and minority fathers rated their infants higher on activity level ($M = 4.93$, $S.D. = .99$; $t(77) = -2.27$, $p < .05$) and distress to limits ($M = 3.69$, $S.D. = .80$; $t(77) = -2.21$, $p < .05$) and lower on soothability ($M = 4.42$, $S.D. = .42$; $t(76) = 2.27$, $p < .05$), and rated themselves higher on depressive symptoms ($M = 1.64$, $S.D. = .47$; $t(12.33) = -2.39$, $p < .05$) than White fathers ($M = 4.38$, $S.D. = .73$; $M = 3.27$, $S.D. = .60$; $M = 4.84$, $S.D. = .59$; $M = 1.31$, $S.D. = .27$, respectively). However, there were no differences between minority and White participants for infant observed fear and observed anger; thus race was not a viable covariate. Simple correlations between family income and parents' education and age and each of the IBQ-R subscales for both mothers and fathers and the observed measures of temperament were calculated. Family income was correlated with mother reported high pleasure ($r(99) = -.25$, $p < .05$), mother reported perceptual sensitivity ($r(99) = -.23$, $p < .05$), mother reported approach ($r(99) = -.26$, $p < .01$), and father reported distress to limits ($r(74) = -.27$, $p < .05$). Mother age was correlated with mother reported smiling and laughter ($r(113) = -.32$, $p < .01$), mother reported high pleasure ($r(113) = -.23$, $p < .05$), mother reported perceptual sensitivity ($r(113) = -.20$, $p < .05$), mother reported approach ($r(113) = -.27$, $p < .01$), and mother reported vocal reactivity ($r(113) = -.36$, $p < .01$). Mother education was correlated with mother reported

activity level ($r(113) = -.20, p < .05$), mother reported smiling and laughter ($r(113) = -.33, p < .01$), mother reported high pleasure ($r(113) = -.29, p < .01$), and mother reported vocal reactivity ($r(113) = -.26, p < .01$). In general, mothers' reports of temperament were related to demographics, whereas fathers' reports were not. Surprisingly, older, more affluent, and higher educated mothers tended to rate their infants lower on the positive dimensions of temperament than did other mothers. However, demographics did not correlate with observed temperament, and thus were not included as covariates.

Table 1: Descriptive statistics

Variables	Mothers $N = 115$			Fathers $N = 79$		
	<i>M</i>	<i>S.D.</i>	Range	<i>M</i>	<i>S.D.</i>	Range
Reported temperament						
Approach	5.36	.90	2.42–7.00	5.11	.85	2.92–7.00
Distress to limitations	3.19	.72	1.36–4.69	3.34	.65	1.75–4.94
Fear	2.22	.80	1.00–5.47	2.40	.82	1.00–4.20
Duration of orienting	4.31	1.05	1.60–7.00	4.23	.87	2.38–6.33
Smile/laughter	5.15	.98	2.60–7.00	5.01	1.03	3.00–6.90
High pleasure	5.60	.61	4.38–7.00	5.84	.78	2.73–7.00
Low pleasure	5.55	.79	3.00–7.00	5.30	.76	3.50–7.00
Soothability	5.07	.60	3.40–6.83	4.78	.59	3.29–6.11
Falling reactivity	5.26	.77	3.15–7.00	5.00	.79	3.15–6.36
Cuddliness	5.78	.66	3.53–6.86	5.52	.69	3.86–6.71
Perceptual sensitivity	4.08	1.00	1.36–7.00	4.11	1.06	1.67–6.33
Sadness	3.14	.85	1.33–5.07	3.12	.89	1.00–5.08
Activity level	4.45	.75	2.47–6.73	4.47	.79	2.69–6.27
Vocal reactivity	5.01	.93	2.09–6.83	4.84	.96	2.92–6.56
Depression	1.42	.35	1.00–3.45	1.36	.33	1.00–2.55
Observed temperament, $N = 98$						
Distress to novelty	.00	.53	-1.04 to 2.32			
Distress to limitations	.00	.79	-.69 to 3.79			

Note. Observed temperament composites are standardized.

Internal consistency reliability

Internal consistency reliability was calculated for each of the IBQ-R subscales for both mothers and fathers. As demonstrated in Table 2, internal reliability for each subscale was adequate ($\alpha > .70$) for both mothers and fathers.

Table 2: Internal reliability coefficients and simple correlations among mother and father reports

Subscale	Chronbach's alphas		<i>N</i> (items)	Mother/father correlations, $n = 78$
	Mother	Father		
Approach	.86	.83	12	.46**
Distress to limitations	.76	.71	16	.33**
Fear	.87	.88	16	.33**
Duration of orienting	.81	.73	12	.32**
Smile/laughter	.84	.83	10	.44**
High pleasure	.76	.78	11	.12
Low pleasure	.84	.80	13	.26*
Soothability	.74	.73	18	.08
Falling reactivity	.82	.81	13	.39**
Cuddliness	.88	.82	17	.17
Perceptual sensitivity	.76	.78	12	.29**
Sadness	.84	.84	14	.32**
Activity level	.71	.77	15	.43**
Vocal reactivity	.81	.82	12	.28*

Note. * $p < .05$, ** $p < .01$.

Gender differences

t-tests based on infant gender demonstrated that fathers rated sons higher than daughters on smiling and laughter ($M = 5.20$, $S.D. = 1.08$ vs. $M = 4.73$, $S.D. = .89$; $t(77) = 2.05$, $p < .05$), high pleasure ($M = 6.01$, $S.D. = .77$ vs. $M = 5.61$, $S.D. = .75$; $t(77) = 2.30$, $p < .05$), low pleasure ($M = 5.49$, $S.D. = .75$ vs. $M = 5.04$, $S.D. = .71$; $t(77) = 2.67$, $p < .01$), soothability ($M = 4.91$, $S.D. = .62$ vs. $M = 4.60$, $S.D. = .51$; $t(76) = 2.30$, $p < .05$), and falling reactivity ($M = 5.20$, $S.D. = .73$ vs. $M = 4.73$, $S.D. = .80$; $t(77) = 2.68$, $p < .01$) than daughters. There were no differences based on infant gender for mother reports of temperament or observed indices of temperament.

Inter-rater reliability

Simple correlations were calculated between parallel mother and father reports to test inter-rater reliability for each of the subscales and are reported in Table 2. All correlations were positive and all were significant except high pleasure, soothability, and cuddliness.

Infant gender and concordance between parent reports

Because prior research has demonstrated that parents may rate their sons and daughters differently and fathers in this sample rated sons significantly more positively than daughters on several dimensions, infant gender was examined as a moderator of concordance between mother and father reports of temperament by utilizing hierarchical multiple regression. Mother reports on the IBQ-R subscale and infant gender were entered into the first step of the regression followed by the interaction of mother report and gender to predict fathers' reports on the parallel temperament dimension. None of the interactions significantly predicted father reports on the IBQ-R subscales, thus infant gender did not moderate the degree to which mothers and fathers agree in their ratings of infant temperament.

Concordance between parent reports and observed temperament

Simple correlations between parent reports on relevant IBQ-R subscales and observed temperament behaviors were conducted to examine concordance of parent reports and observed temperament. As illustrated in Table 3, few correlations between parent reports and observed fear and anger behaviors were significant. Mother's reports of fear correlated positively with both observed fear and observed anger. Father's reports of approach correlated negatively with observed fear. Father's reports of fear did not correlate with observed fear and mother's reports of approach did not correlate with observed fear. Additionally, neither mother nor father reports of distress to limits correlated with observed anger.

Table 3: Simple correlations among observed temperament and mother and father reports of temperament

	Observed fear	Observed anger
Mother IBQ-R approach	-.05	-.05
Mother IBQ-R fear	.22*	.32**
Mother IBQ-R distress to limitations	-.00	.15
Father IBQ-R approach	-.28*	-.12
Father IBQ-R fear	.08	.03
Father IBQ-R distress to limitations	.09	.03

Note. * $p < .05$, ** $p < .01$.

Factors that alter concordance of parent reports and observed temperaments

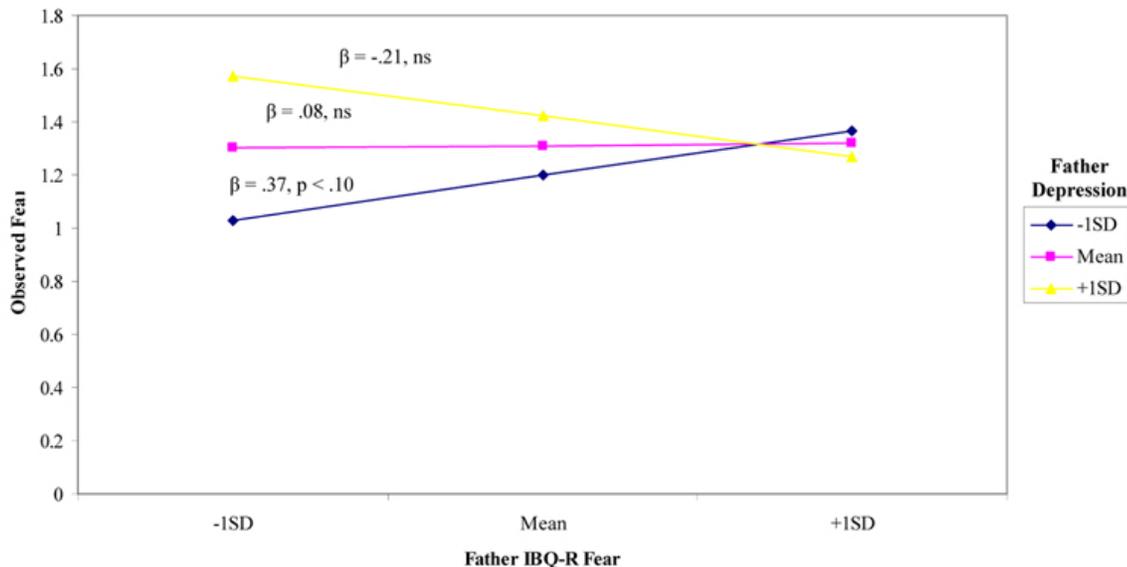
Hierarchical multiple regression was used to test the moderating effects of infant gender and parent depression on the concordance of parent reports and observed temperament. Interaction effects were tested and interpreted using procedures outlined by Aiken and West (1991). To predict observed fear, infant gender, parent depression, parent reported fear, and parent reported approach were entered in the first step of the regression followed by the interaction of depression and reported fear, depression and reported approach, gender and reported fear, and gender and reported approach. As illustrated in Table 4, there was a significant interaction between father depression and father reported fear to predict observed fear. Consistent with the hypothesis, father reports of fear were positively related to observed fear only if fathers were low on depression as illustrated in Fig. 1. Fathers' reports of fear were not related to observed fear if they were average or high on depression. There were no other significant interactions for fathers or mothers to predict observed fear.

Table 4: Hierarchical multiple regressions predicting observed fear

Predictors	Mothers		Fathers	
	β	S.E. (β)	β	S.E. (β)
1. Main effects				
Infant gender	.11	.11	.04	.14
Depressive symptoms	.06	.16	.18	.24
IBQ-R fear	.19***	.08	.02	.09
IBQ-R approach	-.04	.06	-.24***	.08
2. Interactions				
Depression \times IBQ fear	.19	.10	-.24*	.30
Depression \times IBQ approach	-.09	.19	-.22	.22
Gender \times IBQ fear	.32	.15	-.20	.17
Gender \times IBQ approach	-.49	.12	-.30	.19
Adjusted R^2	.12		.12	
F for model	1.73		2.11*	

Note. $N = 98$ for mother analysis, 65 for father analysis. *** $p < .10$, * $p < .05$.

Figure 1: Father depression moderates the concordance between fathers' reports of fear and observed fear.



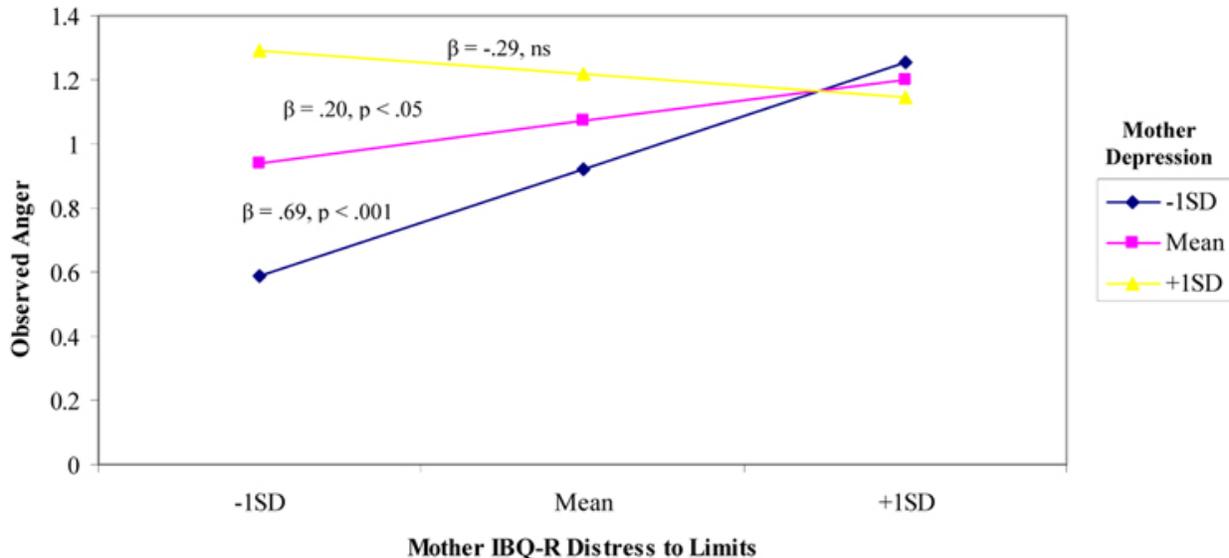
To predict observed anger, infant gender, parent depression, and parent reported distress to limits were entered in the first step of the regression followed by the interaction of depression and reported distress to limits and the interaction of gender and reported distress to limits. As demonstrated in Table 5, there was a significant interaction between mother depression and mother reports of distress to limits to predict observed anger. Consistent with prediction, mother reports of distress to limits were related to observed anger only if mothers were low or average on depression as illustrated in Fig. 2. Mother's reports of distress to limits were not related to observed anger if they were high on depression. There were no other significant interactions for mothers or fathers to predict observed anger.

Table 5: Hierarchical multiple regressions predicting observed anger

Predictors	Mothers		Fathers	
	β	S.E. (β)	β	S.E. (β)
1. Main effects				
Infant gender	.13	.16	.20	.17
Depressive symptoms	.20***	.21	-.18	.28
IBQ-R distress to limits	.16	.11	.04	.14
2. Interactions				
Depression \times IBQ distress to limits	-.32**	.25	.41	.43
Gender \times IBQ distress to limits	.38	.22	.15	.28
Adjusted R^2	.13		.03	
F for model	3.90**		1.47	

Note. $N = 98$ for mother analysis, 66 for father analysis. *** $p < .10$, ** $p < .01$.

Figure 2: Mother depression moderates the concordance between mothers' reports of distress to limits and observed anger.



DISCUSSION

In this study we examined reliability and validity of the IBQ-R at 6 months. The subscales demonstrated good internal consistency reliability and reasonably good inter-rater reliability between mothers and fathers. However, there was limited evidence of convergent validity between parent reports and observed indices, and convergence was moderated by parent depression suggesting that only some parents provide valid assessments of infant temperament.

Reliability

Internal reliability for each subscale of the IBQ-R was adequate and similar to the values reported by Gartstein and Rothbart (2003). Moreover, the internal reliability coefficients were similar for both mothers and fathers indicating the scale structure is comparable for parents of both genders. In addition, inter-rater reliability between mothers and fathers was reasonable for eleven of the fourteen subscales. As expected, parent agreement was stronger with the current sample of families than in Gartstein and Rothbart's sample likely because of the larger sample of secondary caregivers, and the fact that all secondary caregivers in this sample were fathers who likely observe their infants in somewhat more similar contexts to mothers than do other caregivers such as babysitters. Inter-rater reliability was not established for cuddliness, high pleasure or soothability. The lack of agreement on these three subscales may reflect differences in the parenting activities parents engage in or the style in which they interact with their infants. The nonsignificant association for soothability could be a function of differences in parents' abilities to effectively soothe their infants when distressed (Gartstein & Rothbart, 2003). Likewise, differences in the frequency, style, and sensitivity of parents' attempts to be affectionate with their infants could contribute to differences in the infants' response to their overtures explaining the nonsignificant association for cuddliness. The lack of association for high pleasure may be the result of differences in parents' opportunities to observe infants' behaviors in those contexts which pertain to this subscale (e.g., enjoyment when vigorously tickled or tossed into the air) to the extent that fathers tend to engage in more active forms of play with their infants than mothers (Yogman, 1994).

Related to this view, fathers rated sons higher than daughters on smiling and laughter, high pleasure, low pleasure, soothability, and falling reactivity. In contrast, there were no infant gender differences for mother reports of temperament or observed temperament. This may indicate that fathers have a gender bias in how they rate infant temperament such that fathers generally rate their sons higher on more positive dimensions of temperament than they rate daughters. It may also indicate that fathers of male infants interact differently with their children than do fathers of female infants (Power & Parke, 1982; Yogman, 1994) which may elicit differences in infant behaviors. For example, fathers of male infants may engage in vigorous forms of play more frequently than fathers of female infants, thereby eliciting more positive affect from their infants. It is also possible that male infants respond more positively than female infants to these active types of stimulation ([Maccoby, 1988] and [Maccoby, 1998]). This finding was not surprising given previous research which has demonstrated that gender biases exist in how parents rate their infants' temperaments (Diener & Bradshaw, 2002; Polak et al., 2002). Despite these differences in fathers' reports of temperament, infant gender did not moderate agreement between mothers' and fathers' reports of temperament. This finding was inconsistent with our hypothesis as it was expected that agreement would be stronger among parents of female infants than parents of male infants. It is possible that other factors such as the amount of time that parents engage in caregiving activities or the types of activities that parents engage in with their infants may serve as moderators.

Validity

Associations between parallel parent reports and observed indices of temperament were few and small in magnitude. Infants rated as high on fear by mothers and low on approach by fathers displayed greater fear during the observed novelty task demonstrating some convergence as

expected. However, mothers' reports of infant fear correlated also with observed anger. This may indicate that mothers' reports of fear reflect a broader construct of negative emotionality rather than being emotion specific. Alternatively, infants may have demonstrated some fear during the limitations task given it was conducted by a stranger in which case the observational tasks may not be sufficiently emotion specific. Fathers' reports of distress to limits were unrelated to observed infant anger as both a main effect and when considered in conjunction with depressive symptoms. It may be that males have a harder time accurately perceiving frustration cues than females and perhaps father characteristics other than depression, for example, hostility or anger proneness, moderate the degree to which their reports of distress to limits correlate with observed measures of infant anger. Alternatively, infant gender and depressive symptoms may jointly moderate the association between fathers reports of distress to limits and observed anger. Such an effect would necessitate testing three-way interactions which is prohibited in this investigation given the relatively small sample of fathers.

Consistent with previous research and our hypothesis, parent depression moderated the degree of concordance between parent reports and observed indices of temperament (Forman et al., 2003; Leerkes & Crockenberg, 2003). Fathers' reports of fear were positively related to observed fear only if they were low on depression. Likewise, mothers' reports of distress to limitations were positively related to observed anger only if they were low or average on depression. These findings are consistent with prior research which has demonstrated that depressed parents are more likely to misinterpret their infants' signals and rate their infants as more difficult than parents who are not depressed (Atella et al., 2003; Schuetze & Zeskind, 2001). However, contrary to hypothesis, infant gender did not moderate the relationship between parent reports of temperament and observed temperament for mothers or fathers. Thus gender biases held by parents may not explain the lack of concordance between observed and reported temperament. However, to fully examine this hypothesis, future researchers should measure parents' gender ideology as there is likely a great deal of individual difference in this characteristic.

Importantly, the relative lack of concordance between parent reports and observed temperament may reflect problems with the observational context or the degree of match between the observational context and items on the IBQ-R. That is the nature of the observational activities may be more intense than emotionally arousing contexts asked about in the IBQ-R. Furthermore, despite efforts to use comparable measures, the observational tasks overlap with only some items on the IBQ-R subscales. For example, the fear subscale includes items about distress to novel objects, sudden noises, and new people. The novel toy task primarily taps into fear of novel objects. Likewise, the distress to limitations subscale includes items about distress while in confining places or during caretaking activities. The limiting task primarily taps into restriction of movement. However, that parent depression moderated the degree of concordance between parent reports and observed temperament indicates that parent biases explain at least some of the discrepancies.

Limitations and directions for future research

In the current study the assumption was made that observational indices serve as the gold standard for evaluating infant temperament. Observational methods for evaluating temperament may be problematic as they only capture the behaviors infants engage in at a single point in time, and in this case in response to only one type of stressor for each temperament dimension under

consideration. Future research should consider multiple observations of each temperament dimension in multiple contexts to capture stable characteristics of infant temperament and to explore the relative lack of concordance between parent reports and observed temperament in the current study. In addition, the predictive validity of each type of temperament measure must be examined to best determine if one or both approaches are ultimately more useful in predicting family and child outcomes.

ACKNOWLEDGEMENTS

This research was funded by the following grants awarded to the second author: New Faculty Grant and Summer Excellence Award from the Office of Sponsored Programs and seed money from the Human Environmental Sciences Center for Research at The University of North Carolina at Greensboro and the National Institute for Child Health and Human Development (R03HD048691). We are grateful to the childbirth educators who allowed us to enter their classes for recruitment, to the families who generously gave their time to participate in this study, and to Anna Hussey, Kate Seymour, Cate Nixon, and Mary Beth Lee for assistance with data collection and entry.

REFERENCES

- Aiken, L., & West, S. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- Antill, J. K. (1987). Parents' beliefs and values about sex roles, sex differences, and sexuality. *Review of Personality and Social Psychology, 7*, 294–328.
- Atella, L. D., DiPietro, J. A., Smith, B. A., & St. James-Roberts, I. (2003). More than meets the eye: Parental and infant contributors to maternal and paternal reports of early infant difficulty. *Parenting: Science and Practice, 3*, 265–284.
- Bates, J. E. (1980). The concept of difficult temperament. *Merrill-Palmer Quarterly, 26*, 299–319.
- Bates, J. E., & Bayles, K. (1984). Objective and subjective components in mothers' perceptions of their children from age 6 months to three years. *Merrill-Palmer Quarterly, 30*, 111–130.
- Brody, L. R., & Carter, A. (1982). Children's emotional attributions to self versus other: An exploration of an assumption underlying projective techniques. *Journal of Consulting and Clinical Psychology, 50*, 665–671.
- Braungart-Rieker, J.M., & Stifter, C.A. (1996). Infants' response to frustrating situations: Continuity and change in reactivity and regulation. *Child Development, 67*, 1767–1779.
- Calkins, S. D., & Degnan, K.A. (2006). Temperament in early development: Implications for childhood psychopathology. In R. Ammerman (Ed.), *Comprehensive handbook of childhood psychopathology* (pp. 64–84). New York: Wiley.

- Crockenberg, S. C., & Leerkes, E. M. (2003). Parental acceptance, postpartum depression, and maternal sensitivity: Mediating and moderating processes. *Journal of Family Psychology, 17*, 80–93.
- Dave, S., Nazareth, I., Sherr, L., & Senior, R. (2005). The association of parental mood and infant temperament: A pilot study. *British Journal of Developmental Psychology, 23*, 609–621.
- Diener, M. L., & Bradshaw, A. (2002). *Understanding toddler temperament: Associations among temperament, maternal, and child characteristics*. Poster presented at the Biennial Meeting of the International Conference on Infant Studies, Toronto, Canada.
- Donovan, W. L., Leavitt, L. A., & Walsh, R. O. (1998). Conflict and depression predict maternal sensitivity to infant cries. *Infant Behavior & Development, 21*, 505–517.
- Forman, D. R., O'Hara, M. W., Laren, K., Coy, K. C., Gorman, L. L., & Stuart, S. (2003). Infant emotionality: Observational methods and the validity of maternal reports. *Infancy, 4*, 541–565.
- Gartstein, M. A., & Rothbart, M. K. (2003). Studying infant temperament via the Revised Infant Behavior Questionnaire. *Infant Behavior & Development, 26*, 64–86.
- Gill, K. L., & Link, S. D. (2000). Factors affecting concordance between laboratory assessment and maternal perception of infant temperament. In *Poster presented at the Biennial Meeting of the International Society of Infant Studies*.
- Goldsmith, H. H., & Rothbart, M. K. (1996). *The Laboratory Temperament Assessment Battery (Prelocomotor version 3.0)*. Unpublished manuscript. University of Wisconsin, Madison.
- Leerkes, E. M., & Crockenberg, S. C. (2003). The impact of maternal characteristics and sensitivity on the concordance between maternal reports and laboratory observations of infant negative emotionality. *Infancy, 4*, 517–539.
- Maccoby, E. E. (1988). Gender as a social category. *Developmental Psychology, 24*, 755–765.
- Maccoby, E. E. (1998). *The two sexes: Growing apart, coming together*. Cambridge, MA: Harvard University Press.
- Mebert, C. J. (1991). Dimensions of subjectivity in parents' ratings of infant temperament. *Child Development, 62*, 352–361.
- Paulussen-Hoogeboom, M. C., Stams, G. J., Hermanns, J. M., & Peetsma, T. T. (2007). Child negative emotionality and parenting from infancy to preschool: A meta-analytic review. *Developmental Psychology, 43*, 438–453.

- Plant, E. A., Hyde, J. S., Keitner, D., & Devine, P. G. (2000). The gender stereotyping of emotions. *Psychology of Women Quarterly*, *24*, 81–92.
- Polak, C. P., Henderson, H.A., & Fox, N.A. (2002). *Socialization of temperamental anger*. Poster presented at the Biennial Meeting of the International Conference on Infant Studies, Toronto, Canada.
- Power, T. G., & Parke, R. D. (1982). Play as a context for early learning: Lab and home analyses. In I. E. Sigel & L. M. Laosa (Eds.), *The Family as a Learning Environment* (pp. 147–178). New York: Plenum.
- Radloff, J. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*, 385–401.
- Rothbart, M. K. (1981). Measurement of temperament in infancy. *Child Development*, *52*, 569–578.
- Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at 3–7 years: The Children’s Behavior Questionnaire. *Child Development*, *72*, 1394–1408.
- Rothbart, M. K., & Bates, J. E. (1998). Temperament. In N. Eisenberg (Ed.), *Social, emotional, and personality development: Vol. 3 Handbook of child psychology* (pp. 105–176). New York: Wiley.
- Sameroff, A. J., Seifer, R., & Elias, P. K. (1982). Sociocultural variability in infant temperament ratings. *Child Development*, *53*, 164–173.
- Sandnabba, N. K., & Ahlberg, C. (1999). Parents’ attitudes and expectations about children’s cross-gender behavior. *Sex Roles*, *40*, 249–263.
- Schuetze, P., & Zeskind, P. S. (2001). Relationships between women’s depressive symptoms and perceptions of infant distress signals varying in pitch. *Infancy*, *2*, 483–499.
- Siegal, M. (1987). Are sons and daughters treated more differently by fathers than mothers? *Developmental Review*, *7*, 183–209.
- Spitzer, R. L., Endicott, J., & Robins, E. (1978). Research diagnostic criteria: Rationale and reliability. *Archives of General Psychiatry*, *35*, 773–782.
- Thomas, A., Chess, S., & Birch, H. G. (1968). *Temperament and behavior disorders in children*. New York: New York University Press.
- Vaughn, B. E., Bradley, C. F., Joffe, L. S., Seifer, R., & Barglow, P. (1987). Maternal characteristics measured prenatally are predictive of ratings of temperamental difficulty

on the Carey Infant Temperament Questionnaire. *Developmental Psychology*, 23, 152–161.

Yogman, M. W. (1994). Observations on the father–infant relationship. In S. H. Cath, A. R. Gurwitt, & J. M. Ross (Eds.), *Father and child: Developmental and clinical perspectives* (pp. 101–122). Hillsdale: The Analytic Press.