Brief report

# Measuring negative emotionality using the infant behavior questionnaire-revised very short form in a low income, diverse sample

Amy Van Schagen Johnson [a], Esther M. Leerkes [a], Beth A. Reboussin [b],
Stephanie S. Daniel [a], C. Chris Payne [a], Joseph G. Grzywacz [c]

[a] The University of North Carolina at Greensboro, Greensboro, NC, United States
[b] Wake Forest School of Medicine, Winston-Salem, NC, United States
[c] Florida State University, Tallahassee, FL, United States

Infant temperament, particularly negative emotionality, is a frequently studied construct in infancy given its links with later child outcomes and family functioning. For example, heightened negative emotionality in infancy is associated with later problem behaviors in young children such as internalizing and externalizing symptoms (Rothbart & Bates, 2006). In addition, infant negative emotionality is associated with greater parenting stress, depression, and marital difficulties, and with less sensitive parental behavior particularly when other risks are present (Crockenberg & Leerkes, 2003). Therefore, it is important to adequately measure negative emotionality early in life. Negative emotionality has been described as a dimension of temperament that includes the frequency and intensity with which infants experience emotions such as sadness, frustration/anger, fear, and discomfort (Rothbart & Bates, 2006).

Typically, negative emotionality is measured using either trained observer ratings or parent report of child behavior. Specifically, infant behavior during novelty tasks and limiting tasks, believed to elicit fear and frustration, are often videotaped and then later coded for negative emotionality (e.g., Parade & Leerkes, 2008). Although observations have been found effective in measuring negative emotionality, it is a costly and time-consuming process that may be unrealistic in large-scale research. The Infant Behavior Questionnaire Revised (IBQ-R) and related Very Short Form (IBQ-R VSF) are parent-report questionnaires that identify three broad scales including negative emotionality (Gartstein & Rothbart, 2003; Putnam, Helbig, Gartstein, Rothbart, & Leerkes, 2014). The IBQ-R and IBQ-R VSF have been used and validated using predominantly White middle class samples (Putnam et al., 2014). To our knowledge the psychometric properties of the IBQ-R VSF Negative Emotionality scale have yet to be examined using a diverse sample or compared across different ethnic groups. The purpose of this study is to determine if the negative emotionality subscale of the IBQ-R VSF demonstrates measurement invariance and comparable convergent validity with observed indices of negative affect across White and non-White mothers with low income. We focus on the negative emotionality subscale for two reasons: (1) negative emotionality is commonly studied in relation to child and family outcomes and (2) we had observed measures of negative emotionality but not of orienting or positive affect/surgency in our data set.

Participants were 285 low income working mothers and their infants. Eligible participants were women who delivered a live birth within the past 3 months, intended to return to work 27 or more hours per week, and whose household earnings

**Table 1**
Demographics for White and Non-White participants.

|  | White | Non-White |
|---|---|---|
|  | Mean (SD) | Mean (SD) |
| Age | 26.1 (4.9) | 27.6 (5.3) |
| Number of children in household | 1.9 (1.2) | 2.3 (1.3) |
| Educational attainment | *n* (%) | *n* (%) |
| High school, GED, or less | 35 (41%) | 52 (26%) |
| Post high school or certificate | 41 (48%) | 115 (58%) |
| 4 Year degree or more | 9 (11%) | 33 (17%) |
| Marital status |  |  |
|   Currently married | 26 (31%) | 34 (17%) |
|   Living as married | 19 (22%) | 21 (11%) |
|   Single (divorced, widowed, never married) | 40 (47%) | 145 (73%) |
| Child gender |  |  |
|   Male | 40 (47%) | 111 (56%) |
|   Female | 45 (53%) | 89 (45%) |

were within 185% of federal poverty thresholds. The institutions ethics review boards approved the sampling, recruitment, and data collection procedures. Demographic information for White ($n = 85$, 10 of whom were Hispanic) and Non-White ($n = 200$; 182 African American [2 of whom were Hispanic], 3 American Indian or Native Alaska, 1 Asian, 13 multi-racial, and 1 other) mothers is presented in Table 1.

When infants were 3 months old, two project staff members conducted a home visit during which mothers reported on family demographics and completed the IBQ-R VSF (Putnam et al., 2014) on which they rated the frequency with which their infants' engaged in specific behaviors during the past seven days on a seven-point likert scale (1 being Never and 7 being Always). The focus of this report is the 12-item Negative Emotionality subscale which includes items reflecting Sadness, Distress to Limitations, and Fear. Example statements include: "When tired, how often did your baby show distress?" and "When in the presence of several unfamiliar adults, how often did the baby cling to a parent?" In prior research, this subscale has demonstrated internal consistency, test retest reliability, and convergence with observed infant temperament across multiple data sets (Putnam et al., 2014).

Then, mothers and infants were videotaped during a measuring series involving routine caregiving activities and a 4-minute arm restraint. Each mother was asked to undress her infant, leaving only the diaper in place, and to place the infant into a reclining infant car seat that was positioned on a digital scale. Mothers then assisted the data collector in obtaining the child's heel-crown length by placing the infant into an infantometer (Perspective Enterprises, Model No. PE-RILB-LTWT) and holding the child in place to enable assessment to the nearest .1 cm. Weight and height were each measured at least twice to ensure accuracy. This task is ideal for rating infant affect as it can be viewed as a mildly stressful experience with high ecological validity as the experience of being dressed/undressed and repositioned by caregivers is typical.

Next, mothers were asked to buckle their infant into a car seat that was positioned on the floor. Mothers were instructed to sit on the floor to the infant's right, and a small bag of age appropriate toys was placed within the mother's reach. The research assistant knelt behind the infant seat and gently held the infant's forearms immobile for 4 min. During the first minute, the mother was instructed to remain neutral and uninvolved unless she wanted to end the activity. Then the research assistant signaled the mother that she could interact with her infant as she pleased for the remaining 3 min. This task, designed to elicit infant frustration, has been used in prior research of this type (Parade & Leerkes, 2008).

For both tasks, infant negative affect was coded in 15-s epochs on a 4 point scale (0 = no distress; 3 = high distress) yielding measures of peak distress, proportion of epochs distressed, and average distress for each task. Task start time and the time at which the infant first displayed distress were noted to calculate latency to distress. Inter-rater reliability was established for each behavior based on 37 double coded videos; intraclass correlation coefficients ranged from .85 to .99. Scores within a task were highly correlated (*rs* .68–.99 absolute value), and were standardized and averaged with latency to distress reverse scored yielding measures in which high scores reflect greater observed distress in each task ($\alpha$ = .96, for both caregiving and arm restraint task).

Prior to conducting the primary analyses, we conducted *t*-tests to examine possible differences in mother reported and observed infant negative emotionality based on infant gender. Males were rated as marginally higher on IBQ-R VSF negative emotionality (*M* = 1.31, SD = 11.63) and observed distress during the arm restraint task (*M* = .52, SD = 4.77) compared to females (*M* = −1.48, SD = 11.78, *t*(258) = −1.92, *p* = .06, and *M* = −.54, SD = 4.39, *t*(282) = −1.94, *p* = .05), for IBQ-R VSF and observed, respectively. Thus, we controlled for gender in our regression analyses.

Multiple group confirmatory factor analysis was carried out in several stages to examine whether the negative emotionality subscale of the IBQ-R VSF demonstrates measurement invariance across White and non-White mothers with low income. Testing for measurement invariance proceeds by comparing a series of models that define more and more stringent equality constraints. First, configural invariance was tested by fitting a model in which all items load on a single negative emotionality dimension; herein referred to as the baseline model. Error terms were correlated for items from the same subscale on the original IBQ-R consistent with Putnam et al. (2014). The magnitude of all parameters was allowed to vary