Measuring negative emotionality using the Infant Behavior Questionnaire-Revised Very Short Form in a low income, diverse sample

By: Amy Van Shagen Johnson, Esther M. Leerkes, Beth A. Reboussin, Stephanie S. Daniel, C. Chris Payne, and Joseph G. Grzywacz

Van Shagen Johnson, A., Leerkes, E.M., Reboussin, B.A., M., Daniel, S.S., Payne, C.C., & Grzywacz, J.G. (2016). Measuring negative emotionality using the Infant Behavior Questionnaire-Revised Very Short Form in a low income, diverse sample. *Infant Behavior and Development, 42,* 100-103.

Made available courtesy of Elsevier: https://doi.org/10.1016/j.infbeh.2015.11.004

 **This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.**

*****© Elsevier. Reprinted with permission. This version of the document is not the version of record. Figures and/or pictures may be missing from this format of the document. *****

Abstract:

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).

Keywords: infant temperament | parental stress | negative infant emotionality | infant distress

Article:

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 and Rothbart, 2003, Putnam et al., 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 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.

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

	White Mean (SD)	Non-White 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%)

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 (*r*s .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 (α = .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 between groups. Following Cheung and Rensvold (2002), the root mean square error of approximation (RMSEA) was used to assess configural variance. The comparative fit index (CFI) was used as a supplemental index. Adequate model fit is indicated by RMSEA values from .06 to .08 and CFI values between .90 and .95 (Hu & Bentler, 1999). Constraining all factor loadings to be equal across groups and comparing to the baseline model then tested metric invariance. Metric invariance exists if the fit of the model is not substantially worse than the fit of the baseline model. Next, the factor loadings and item intercepts were constrained to be equal to test scalar invariance and compared against the metric invariance model that only constrains the loadings. Finally, restricting the loadings, item intercepts and error variances to be equal and then comparing to the scalar invariance mode tests error invariance. Because the sequence of models is nested, chi-square difference ($\Delta\chi^2$) tests are used for model comparisons. Given that chi-square difference tests are sensitive to sample size, we also relied on the difference in CFI (ΔCFI) as recommended by Cheung and Rensvold (2002). A cutpoint of ΔCFI < .01 is often used as an indicator that the constrained model does not result in a substantial decrease in model fit compared to a less constrained model. All analyses were performed in Mplus Version 7.11. Next, multiple linear regression models were used to examine if the convergent validity of the IBQ-R VSF with observed affect in the caregiving and the arm restraint tasks varied as a function of maternal race. Infant gender, a covariate, maternal race, and mother reports of negative emotionality on the IBQR-VSF (centered) were entered in the first step (Model 1), and the race by IBQ-R VSF product term was entered in the second step (Model 2).

The overall RMSEA value for the baseline configural invariance model was .065 with a 90% confidence interval between .042 and .087. The CFI value was .940. These results suggest that model fit is adequate and configural variance can be assumed across racial groups. Although the change in CFI was slightly greater than .01, the test of metric invariance suggests that the model fit did not get significantly worse when all of the factor loadings were constrained to be equal (ΔCFI = .012, $\Delta\chi^2 = 19.0$, df = 11, $p = .061$). In the next step when the intercepts were constrained to be equal, the model fit got significantly worse suggesting a lack of full scalar invariance (ΔCFI = .016, $\Delta\chi^2 = 23.9$, df = 12, $p = .021$). We then reviewed modification indices, which are estimates of how much the chi-square for the model would be reduced if a single parameter is made free, to identify which specific item intercepts should be freed across groups to help improve model fit. A modification index greater than 3.84 corresponds to a change significant at the .05 level. This approach indicated that allowing two item intercepts (items 4 "When introduced to an unfamiliar adult, how often did the baby cling to a parent?" and 29 "When you were busy with another activity, and your baby was not able to get your attention,

how often did s/he cry?") to vary freely across groups improved model fit, demonstrating partial scalar invariance (ΔCFI = .002, $\Delta\chi^2$ = 11.8, df = 10, p = .299). The last step was to constrain the error variances to be equal. The chi-square difference test was not statistically significant ($\Delta\chi^2$ = 15.2, df = 12, p = .231) and the change in the CFI was less than .01 suggesting that full error variance was satisfied; constraining the error variances to be the same across racial groups did not result in a significant decrease in model fit. In sum, the results demonstrate that the IBQ-R VSF negative emotionality scale is primarily invariant across White and Non-White mothers.

Simple correlations demonstrated that mothers' reports of negative emotionality on the IBQ-R VSF were positively associated with observed infant affect during the measuring series (r = .23) and arm restraint (r = .14) tasks (p < .05). The association between mother reported negative emotionality and observed infant affect remained significant in the regressions over and above infant gender and maternal race (see Table 2). These relationships did not vary by race as indicated by non-significant interactions between race and negative emotionality (see Model 2).

Table 2. Linear regression of IBQ-R VSF on observed infant affect adjusting for race.

	Measuring series[1]				Arm restraint[2]			
	B	SE (*B*)	*β*	*p*-Value	*B*	SE (*B*)	*β*	*p*-Value
Model 1								
Gender	.42	.56	.05	.46	1.20	.57	.13	.04
Race	.65	.61	.06	.29	.04	.63	.00	.95
IBQ-R VSF	.08	.02	.22	.01	.05	.02	.12	.05
Model 2								
Race × IBQ-R VSF	.05	.05	.10	.39	.05	.05	.11	.34

[1] n = 260 due to item missingness.
[2] n = 259 due to item missingness.

Findings indicate that mother reported negative emotionality correlated with observed distress during the measuring series and arm restraint task. Although the association was small, this is comparable to results from previous studies (Parade & Leerkes, 2008). To our knowledge, we are the first to demonstrate that this measure of negative infant emotionality is invariant, both in terms of factor structure and convergent validity with observed distress, across minority and non-minority mothers. This is important because it justifies the use of the IBQ-R VSF in racially diverse samples.

In conclusion, the IBQ-R VSF provides a time efficient and cost-effective way to collect information about temperament from caregivers who have the opportunity to observe children over long periods of time and in multiple contexts (Rothbart & Bates, 2006). Our results indicate this is a valid measure of negative emotionality in racially diverse and low income samples. Future research should examine the psychometric properties of the IBQ-R VSF in older infants as it was intended for use with three to 12 month olds and with fathers and other caregivers. In addition, efforts should be made to examine the invariance of the other two subscales (positive affect/surgency and orienting/regulatory capacity).

Acknowledgments
We are grateful to the participating families and project staff. This project was supported by the Eunice Kennedy Shriver National Institute Of Child Health & Human Development or the National Institutes of Health NIH Grant number: R01-HD061010. A fuller report will be provided upon request.

References

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. Structural Equation Modeling: 9., (2), 233–255. http://dx.doi.org/10.1207/S15328007SEM0902_5

Crockenberg, S., & Leerkes, E. M. (2003). Negative infant emotionality and the development of family relationships in infancy and early childhood. In A. C. Crouter, & A. Booth (Eds.), Children's influence on family dynamics: The neglected side of family relationships (pp. 57–78). Mahwah, NJ: Erlbaum.

Gartstein, M., & Rothbart, M. K. (2003). Studying infant temperament via the Revised Infant Behavior Questionnaire. Infant Behavior and Development: 26., (1), 64–86.

Hu, L., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. Structural Equation Modeling: 6., (1), 1–55.

Parade, S. H., & Leerkes, E. M. (2008). The reliability and validity of the infant behavior questionnaire—Revised. Infant Behavior & Development: 31., 637–646. http://dx.doi.org/10.1016/j.infbeh.2008.07.009

Putnam, S. P., Helbig, A. L., Gartstein, M. A., Rothbart, M. K., & Leerkes, E. (2014). Development and assessment of short and very short forms of the infant behavior questionnaire—Revised. Journal of Personality Assessment: 96., (4), 445–458. http://dx.doi.org/10.1080/00223891.2013.841171

Rothbart, M. K., & Bates, J. E. (2006). Temperament. In N. Eisenberg, & W. Damon (Eds.), Handbook of child psychology: Vol. 3 Social, emotional, and personality development (6th ed., pp. 99–166). New York, NY: Wiley.