

RESEARCH

Open Access



# Standardized observation of temperament in Lebanese toddlers using the laboratory temperament assessment battery (Lab-TAB)

W. Shamseddeen<sup>1,2</sup>, B. Safieddine<sup>1</sup>, H. Mohsen<sup>1</sup> and L. Akoury Dirani<sup>1\*</sup>

## Abstract

**Background** Temperament is the difference between individuals' emotional and behavioral responses to diverse external events. It is a complex interplay between genetic and environmental factors. Hence, the need to assess temperament objectively and better understand its impact on developmental and interpersonal outcomes. Measuring temperament in early childhood can be challenging since parents will report their subjective perceptions about their toddlers. While surveys are quick instruments that require less clinical involvement, standardized laboratory assessments secure a relatively high level of objective observation. Since no published studies were conducted in Arab countries, the current research focuses on examining temperament in a sample of twenty mother-toddler dyads using the Laboratory Temperament Assessment Battery (Lab-TAB) locomotor version.

Interrater reliability and validity were assessed. Higher-order temperament components were determined by principal component analysis. *T* test and one-way ANOVA examined the association between demographics and temperament components.

**Results** The retained variables ranged between fair ( $> 0.43$ ) and good ( $< 0.98$ ) for all Lab-TAB episodes. Three higher-order temperament components were obtained. Age was significantly negatively correlated with Lab-TAB Fear dimension,  $r = -.47, p > .05$ , and Lab-TAB temperament component 3,  $r = .45, p > .05$ . Male toddlers ( $M = .55, SD = 1.055$ ) had significantly higher levels of temperament component 3 compared to female ( $M = -.45, SD = .718$ ),  $t(18) = 2.52, p < .05$ . There was a significant effect of time spent with mother on temperament component 3,  $F(2,17) = 7.01, p < .05$ .

**Conclusion** After exploring the temperament factor structure, we found that the Lab-TAB locomotor version was a valid tool to be used to observe temperament in toddlers living in Lebanon, a Middle Eastern culture. Some gender significant differences would deserve deeper exploration in future research. A replication of this study would also strengthen its findings.

**Keywords** The Laboratory Temperament Assessment Battery (Lab-TAB) clinical observation, Temperament, Toddlers, Lebanon

## Background

Temperament is the difference between individuals' emotional and behavioral responses to diverse external events. It is a complex interplay between genetic and environmental factors [10, 34] leading to resilience or vulnerabilities throughout a life course. Temperament's expression in early childhood is particularly sensitive to parenting, which in turn is influenced by parents' mental health, culture, and socio-economic status.

\*Correspondence:

L. Akoury Dirani  
la55@aub.edu.lb

<sup>1</sup> Department of Psychiatry, American University of Beirut, Beirut, Lebanon

<sup>2</sup> Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA

Rich scientific literature supports these assumptions. Parenting, being positive or negative, impacts the development of the child's social behavior [8]. Cross-cultural differences in parents' perceptions of their children's temperament have been found in a sample of USA and Finns [14] where Americans had higher rates of negative effects as compared to Finns. Socioeconomic factors also have an impact on early childhood temperament and adult psychopathology [27]. Finally, De Pauw and Merielde [9] demonstrated the role of temperament in the likelihood of developing anxiety, depression, ADHD, and proactive and reactive antisocial behavior. Thus, it is crucial for parents and caregivers to better understand the child's baseline temperament, hence providing them with appropriate socioemotional care.

There is a variety of measurements used to understand temperament and its relation to developmental outcomes. These measurements include parents' and teachers' questionnaires, home-based observations, and laboratory observations. Questionnaires are user-friendly and cost-effective; yet, as the majority of self-reports, they are subjective and may be biased. The presence of a stranger doing the home-based observations may alter the child's and the parents' behavior. Standardized laboratory assessments secure a relatively high level of objective observation.

Studies by M.K Rothbart and colleagues became a cornerstone in defining the concept of temperament, and designing measurements of temperament from infancy to adulthood, namely the Infant Behavior Questionnaire-Revised (IBQ-R) [16], the Early Childhood Behavior Questionnaire (ECBQ) [30], the Childhood Behavior Questionnaire (CBQ) [33], the Early Adolescence Temperament Questionnaire (EATQ) [5], and the Adulthood Temperament Questionnaire (ATQ) [11], home-based observations [32], and structured laboratory observations, Lab-TAB prelocomotor [21], locomotor [20], toddler [4], preschool [18], and middle childhood version [17].

To the authors' knowledge, the Lab-TAB is the only laboratory-structured protocol measuring temperament throughout childhood. In addition, Lab-TAB has been used to study the interaction between parent behaviors and child temperament as risk factors for obesity in children [1], to study mothers' perceptions of their child's temperament [24], and to examine "fear" as a predictor for anxiety development in early school years [2, 3]. We found two validation studies of Lab-TAB in two specific cultural contexts, Portuguese [12] and Ethiopian [23]. Since no published studies were conducted in Arab countries, the current research focuses on examining temperament in a group of typically developed Lebanese toddlers using Lab-TAB.

## Methods

This study is part of a larger one aiming at validating the Early Childhood Behavioral Questionnaire (ECBR)-Short version [31], March). One hundred and sixty-two mothers were recruited for the validation study and 20 agreed to participate in the standardized clinical observation using the Locomotor version of the Lab-TAB [20]. Mothers and their children were recruited from public primary health care centers, nurseries, business and academic institutions, and private pediatric clinics. The inclusion criterion was being an adult mother (above 18 years old) of typically developed toddlers. The approval of the Institution Review Board of the American University of Beirut was secured prior to implementing the study.

The Lab-TAB protocol-locomotor version was tested using interrater reliability. The validity of the variables in describing an episode was explored. Then the validity of episodes in describing a given dimension was studied. Finally, the dimensions were grouped into temperamental characteristics.

## Procedure

Consenting mothers filled out a demographic questionnaire prior to the experiment. During the laboratory visit, the toddler was accompanied by the mother, an experimenter, and a videographer. The parent was present in the room with the toddler throughout the experiment and was instructed to remain as uninvolved as possible if the toddler tried to engage their attention unless instructed otherwise. The same experiment protocol was used with all children. Episodes including dialogue and verbal prompts were translated and adapted to Arabic. All laboratory episodes were videotaped for later scoring by three independent raters.

## Measure

### *The Laboratory Temperament Assessment Battery-locomotor version*

The Laboratory Temperament Assessment Battery (Lab-TAB; [19]) is a group of standardized instruments for laboratory assessment of temperament ranging from infancy to middle childhood. The Lab-TAB Locomotor version is designed to elicit infant/toddler emotional and behavioral responses to stimuli across five dimensions of toddler temperament: fear, anger, pleasure, interest/persistence, and activity level. It consists of 20 episodes that are designed for infants that have started to crawl [22]. Each episode consists of scored variables to measure toddler responses to stimuli. For our investigation, two episodes per temperament dimension were chosen similar to what has been done in the Ethiopian study [23]. Episodes were chosen based on ease of execution, cultural relevance,

affordability, and feasibility. A complete description of procedures and coding is found in the Lab-TAB Locomotor manual [22]. The Lab-TAB coding system for toddler emotions relies on facial expressions, bodily movements, and vocalizations. The Lab-TAB has not yet been validated for use in Lebanon.

### Statistical analysis

Before initiating the statistical analysis, data reduction for the Lab-TAB episode and Dimension composites were performed. Data were entered and analyzed using IBM SPSS Statistics 24.

Firstly, interrater reliability was assessed by computing an intra-class correlation coefficient (ICC). Secondly, factor analyses using principal component analysis were conducted to examine the validity of a group of variables in describing a particular episode. After excluding variables with low interrater reliability and lack of validity, the episode composite score was calculated.

Factor analysis was conducted to determine episode loading for each temperament dimension of interest, and then scores were combined across episodes to create an overall score for each temperament dimension. The loadings of episodes on temperament generated from this factor analysis were used to calculate the dimensions scores from their corresponding episodes' scores. For example, to obtain a composite score for the activity level dimension, the mean score of the previously mentioned episode composite scores was calculated. The resulting temperament dimension composite scores were fear, anger and sadness, pleasure, interest/persistence, and activity level.

Moreover, the five obtained temperament dimension composite scores were entered into a factor analysis to determine higher-order temperament components. Finally, independent sample t-test and one-way ANOVA were used to examine which categorical demographic variables (e.g., gender or time spent with mother) were significantly associated with the temperament components. The Pearson correlation coefficient was used to examine the association between temperament and continuous variables such as age. Continuous variables are presented with means and standard deviations, while categorical variables are presented as numbers and frequencies. *P* values less than 0.05 were considered to be statistically significant.

### Data reduction for the LAB-TAB episode and dimension composites

The standardized procedure for scoring Lab-TAB produces raw scores for each epoch across each trial within each episode of the manual. Our data reduction strategy to derive temperament composite scores was adapted from Planalp et al. [29]. The first step of this strategy was

to create response parameters for behaviors across each episode. For instance, in the task orientation episode, the toddler received a raw score for facial interest on each epoch ( $n=6$ ) across each interval ( $n=3$ ), resulting in 18 scores. We averaged facial interest across the 18 epochs and used this as a mean facial interest score for Task Orientation. Calculating the mean level for each behavior helps to quantify the toddler's average reactivity across episodes. Additionally, we identified the peak level of facial interest for the episode, which is the highest score across all 18 epochs. Compared to mean levels, peak levels depict the toddler's tendency to have extreme behavioral responses to stimuli more accurately. Latency scores were not calculated because they were not found to add significant variation in episode-level scoring in previous studies [29].

## Results

### Participants

Mothers and their toddlers ( $N=20$  toddlers) were assessed during a laboratory visit. The toddlers were between 22 and 36 months of age ( $M=29.4$ ,  $SD=4.58$ ). Approximately half of the toddlers ( $n=11$ , 55%) were females. Sixty percent of the mothers were between 30 and 49 years of age. All the mothers were married. Most of the mothers had completed graduate studies (65%), employed (60%), and with a monthly family income above \$2000 (70%). Half of the mothers spent between 3 and 5 h with their toddlers daily. Only 4% of the families sought mental health services for one of their members (refer to Table 1).

### Raters and interrater reliability

Raters included three staff members with master's degrees in psychology, two of whom were doing clinical training in child psychology. Raters were trained by a child clinical psychologist. One rater scored episodes for all participants, while each of the other two other raters scored episodes for half of the participants. Interrater agreement on episode variables is described below.

Interrater reliability for episode coding was assessed by computing an intra-class correlation coefficient (ICC), using a two-way random, absolute agreement, average-measures ICC [28] to assess the degree that coders provided consistency in their ratings of the Lab-TAB episode variables. ICC was calculated for raters 1 and 2 as well as raters 1 and 3, and then average ICC values were calculated for each Lab-TAB episode variable. Interrater reliability was examined for each episode variable. Those with acceptable ICC were retained for analysis. Mean and peak scores for retained episode variables were combined to create episode-level scores. According to the commonly cited cutoffs by Cicchetti [6], variables with ICC

**Table 1** Demographic variables

Variable	Frequency	Percentage
Parents' age		
18–29 years	8	40
30–49 years	12	60
Educational level		
Undergraduate	7	35
Graduate	13	65
Work status		
No work	8	40
Work	12	60
Child gender		
Male	9	45
Female	11	55
Time to spend with child		
Between 1 and 3 h	3	15
Between 3 and 5 h	10	50
All day	7	35
Family seeking mental health services		
No	16	80
Yes	4	20

values above 0.40 were retained for further analysis. In the current study, the retained variables ranged between fair ( $>0.43$ ) and good ( $<0.98$ ) for all Lab-TAB episodes. Variables with ICC values below 0.40 or 0 variance in interrater reliability were excluded and not utilized further.

After examining reliability and validity as detailed below, mean and peak scores were combined across trials to create an overall episode score. Then, scores across episodes were combined to create overall scores for temperament dimensions. For instance, Task Orientation and Person Interest both assess toddler interest/persistence, so the episode composite scores of those episodes were combined to obtain an overall Interest/Persistence composite.

#### Validity of variables in describing an episode

It was also important to examine the validity of grouping episode-level variables (e.g., Mean and peak distress vocalizations, mean intensity of bodily fear, and mean and peak presence of startle response) before calculating episode composite scores (e.g., fear due to unpredictable mechanical toy) in our sample. Therefore, an exploratory factor analysis was conducted using a principal components analysis (PCA) with Oblimin rotation used on the variables to determine the variable loading per episode. To note, varimax rotation was used for the cognitive assimilation game because the episode-level variables were not correlated. Variables that loaded on one factor

were included in computing the composite score of that episode, while the other variables were excluded.

#### Retained variables based on interrater reliability and validity

##### *Fear: Masks*

Interrater reliability for Masks ranged from  $r=0.43$  to 0.93. No variables were excluded at this level. Mean and peak values were combined. An exploratory factor analysis was conducted on the variables: intensity of facial fear, intensity of distress vocalizations, intensity of bodily fear, and intensity of escape behaviors. The factor analysis yielded a two-factor solution for this episode. The first factor was chosen because it had higher eigenvalues. Hence, the intensity of facial fear, the intensity of distress vocalizations, and the intensity of escape behaviors were retained for calculating episode composite scores. Intensity of bodily fear was excluded from further analysis because its factor loading was poor on the chosen factor.

##### *Fear: unpredictable mechanical toy*

Interrater reliability for unpredictable mechanical toy ranged from  $r=-0.03$  to 0.91. Variables excluded based on low interrater reliability scores were intensity of distress vocalizations (mean and peak), intensity of bodily fear (mean), and presence of startle response (mean and peak). Mean and peak values were combined. An exploratory factor analysis was conducted on the variables intensity of facial fear, intensity of bodily fear, and intensity of escape behaviors. All these variables were retained for calculating the episode composite score.

##### *Anger and sadness: car seat restraint*

Interrater reliability for car seat restraint ranged from  $r=-0.09$  to 0.90. Variables that were excluded based on low interrater reliability scores were facial sadness (mean and peak) and body struggle (mean and peak). Mean and peak values were combined. An exploratory factor analysis was conducted on the variables of facial anger, distress vocalizations, and body sadness. All these variables were retained for calculating the episode composite score.

##### *Anger and sadness: gentle arm restraint*

Interrater reliability for gentle arm restraint ranged from  $r=-0.30$  to 0.26. To note, there was zero variance between raters 1 and 2 on most variables. Therefore, variables of this episode were excluded from further analyses.

**Pleasure: puppets**

Interrater reliability for Puppets ranged from  $r = -0.13$  to 0.93. Variables that were excluded based on low interrater reliability scores are laughter (mean and peak) and positive motor acts (peak). Mean and peak values were combined. An exploratory factor analysis was conducted on the variables: intensity of smiling, positive vocalizations, and positive motor acts. These variables were retained for calculating the episode composite score.

**Pleasure: cognitive assimilation game**

Interrater reliability for the cognitive assimilation game ranged from  $r = 0$  to 0.87. Variables that were excluded based on low interrater reliability scores were laughter (mean and peak) and positive motor acts (mean). Mean and peak values were combined. An exploratory factor analysis was conducted using a principal components analysis (PCA) with varimax rotation on the variables smiling, positive vocalizations, and positive motor acts. These variables were retained for calculating the episode composite score.

**Interest/persistence: task orientation**

Interrater reliability for task orientation ranged from  $r = 0$  to 0.98. Variables that were excluded based on low interrater reliability scores were intensity of facial interest (peak), duration of looking (peak), and manipulation of stimuli (peak). Mean and peak values were combined. An exploratory factor analysis was conducted on the variables: intensity of facial interest, duration of looking, and manipulation of stimuli, and these variables were retained for calculating episode composite score.

**Interest/persistence: person interest**

Interrater reliability for Person interest ranged from  $r = 0.71$  to 0.98. No variables were excluded at this level. Mean and peak values were combined. An exploratory factor analysis was conducted on the variables: intensity of facial interest, duration of looking, and vocalizations about the experimenter. All of these variables were retained for further analysis.

**Activity level: free play**

Interrater reliability for free play ranged from  $r = 0.45$  to 0.82. No variables were excluded at this level. Mean and peak values were combined. An exploratory factor analysis was conducted on the variables: intensity of manipulation, bouts of play, intensity of movement, and changes in body position. The factor analysis yielded a two-factor solution, and both factors had similar

eigenvalues. However, variables in the second factor had lower interrater reliability. Therefore, intensity of manipulation and bouts of play were excluded. Variables that were retained for calculating the episode composite score were intensity of movement and changes in body position.

**Activity level: peg/shape manipulation**

Interrater reliability for peg/shape manipulation ranged from  $r = 0.61$  to 0.96. No variables were excluded at this level. Mean and peak values were combined. An exploratory factor analysis was conducted on the variables: number of pegs and intensity of activity. These variables were retained for calculating the episode composite score.

**Validity of episodes in describing a dimension**

Similar to examining the validity of grouping certain variables into the corresponding episode, it was also essential to examine the validity of grouping certain episodes into the corresponding dimension in this sample. Exploratory factor analysis was conducted using a principal components analysis (PCA) with varimax rotation on the episode variables. All episodes loaded as expected on their corresponding temperament dimensions. For instance, the episode variable composites for free play and peg/shape manipulation loaded on the activity level temperament dimension (refer to Table 2).

**Observed toddler temperament**

The five obtained temperament dimension composite scores, namely fear, anger and sadness, pleasure, interest/persistence, and activity level were entered in an exploratory factor analysis using a principal components analysis (PCA) with varimax rotation to determine higher-order temperament components. The obtained results consist of three higher-order temperament components (refer to Table 3). The first component includes pleasure, anger, and sadness. The second component includes interest

**Table 2** Rotated component matrix of Lab-TAB temperament dimensions

Variable	Component		
	1	2	3
Pleasure	.73	.48	.26
Interest/persistence	-.06	.94	-.13
Activity level	.19	-.03	.75
Fear	.21	.06	-.71
Anger and sadness	.87	-.25	-.17

Extraction method: principal component analysis

Rotation method: varimax with Kaiser normalization

Rotation converged in 7 iterations

**Table 3** Correlations between Lab-TAB temperament dimensions

	1	2	3	4	5	6	7	8
1. Pleasure	1							
2. Interest/persistence	.25	1						
3. Activity level	.19	-.11	1					
4. Fear	-.03	.09	-.10	1				
5. Anger and sadness	.36	-.15	-.02	.10	1			
6. Temperament component 1	.73**	-.06	.19	.21	.87**	1		
7. Temperament component 2	.48*	.94**	-.03	.06	-.25	.00	1	
8. Temperament component 3	.26	-.13	.75**	-.71**	-.17	.00	.00	1

\*  $p < .05$ \*\*  $p < .001$ 

and pleasure. The third component includes activity level and fear (negative loading).

#### Variables correlated with temperament

Age was significantly negatively correlated with Lab-TAB Fear dimension,  $r = -0.47$ ,  $p > 0.05$ , and Lab-TAB temperament component 3,  $r = 0.45$ ,  $p > 0.05$ . *T* tests were conducted to examine group differences in toddler temperament. Male toddlers ( $M = 0.55$ ,  $SD = 1.055$ ) had significantly higher levels of temperament component 3 compared to female toddlers ( $M = -0.45$ ,  $SD = 0.718$ ),  $t(18) = 2.52$ ,  $p < 0.05$ . This indicates that female toddlers were more likely to be fearful and less active than male toddlers (refer to Table 4). There were no significant gender differences in toddler temperament on the episode and dimension levels. There were no significant differences in toddler temperament based on mother employment status, mother education, monthly household income, or seeking mental health services for family members.

Moreover, analyses of variance were conducted to examine group differences in toddler temperament. There was a significant effect of time spent with the mother on temperament component 3,  $F(2,17) = 7.01$ ,  $p < 0.05$ . Games-Howell post hoc tests revealed that toddlers who spent between 1 and 3 h with their mothers ( $M = 1.56$ ,  $SD = 0.57$ ) had higher levels of temperament component 3 compared to toddlers who spent between 3 and 5 h ( $M = -0.32$ ,  $SD = 0.76$ ) or all day with their mothers ( $M = -0.22$ ,  $SD = 0.88$ ). No significant effects were noted for time spent with relative or monthly household income.

#### Discussion

This study aimed at validating the Lab-TAB locomotor version to Lebanese toddlers. The toddlers' observed behavior using the Lab-TAB locomotor episodes was carefully analyzed. We examined interrater reliability

**Table 4** Standardized means of Lab-TAB temperament dimensions by gender

	Males ( $n = 9$ )	Females ( $n = 11$ )
Episodes		
Pleasure-puppets	.41	-.33
Pleasure-cognitive assimilation game	.19	-.16
Interest/persistence-task orientation	-.18	.15
Interest/persistence-person interest	.09	-.07
Activity level-peg/shape manipulation	.19	-.15
Activity level-free play	.19	-.15
Fear-unpredictable mechanical toy	-.44	.36
Fear-masks	.27	.22
Anger and sadness-car seat restraint	-.04	.03
Temperament dimensions		
Pleasure	.41	-.34
Interest/persistence	-.07	.06
Activity	.30	-.25
Fear	-.42	.35
Anger and sadness	-.04	.03
Higher-order temperament components		
Temperament component 1	.15	-.12
Temperament component 2	.09	-.07
Temperament component 3	.55	-.45

and validity of episodes and dimensions. After exploring the temperament factor structure, it was found that the Lab-TAB locomotor version was a valid tool to be used to observe temperament in toddlers living in Lebanon. The validated protocol included the 5 temperament dimensions, using 10 episodes instead of the 20 episodes of the original protocol. Each of the fear, pleasure, interest/persistence, and activity level dimensions included

2 episodes, while the anger and sadness dimensions retained only one episode. The duration of the laboratory assessment was 30 to 40 min, which is shorter than the original version.

Three dimensions/factors were identified. The first included pleasure, anger and sadness and we choose to call it “emotional valence” since it associates the emotional expression with a given situation. The second component includes interest and pleasure, and we choose to keep the name “persistence”. The third component includes activity level and fear (negative loading), and we choose to call it “venturesome” since it associates activity level positively loaded with Fear negatively loaded. This last component showed a significant difference between genders where male toddlers showed higher levels of activity and less fear than females. This finding is in line with a large body of research [7, 13, 26]. Gandour [15] showed that activity level although being an individual predisposition is encouraged or refrained by the mother’s stimulation. In our sample, the more toddler spend time with their mothers, the less active and fearful they are. This finding is not supported by evidence from other studies and requires replicated in larger samples and more in-depth analysis. In fact, Hsin and Felfe [25] demonstrated that the time the parent spends with the young child has no impact if not associated with high-quality relationships.

Despite the numerous advantages of the Lab-TAB, some limitations should be noted. First, the administration of the Lab-TAB is time-consuming and expensive. Second, the hesitation of parents to participate in a laboratory assessment due to a lack of familiarity may have contributed to the low sample size. Third, toddlers may react differently in a novel laboratory setting as compared to their natural environment. Considering that conditions in laboratory assessments can impact results, we tried to minimize the effects of examiner and rater differences and toddler state.

## Conclusion

We have explored the temperament factor structure of the Lab-TAB locomotor version and found that this was a valid tool to be used to observe temperament in toddlers living in Lebanon, a Middle Eastern culture. Some gender significant differences would deserve deeper exploration in future research. A replication of this study would also strengthen its findings.

## Authors’ contributions

LAD principal investigator applied for the grant and received it, designed the research and oversee its implementation, data collection and analysis, and the write up of the manuscript.

WS conducted the statistical analysis and contributed to the write up of the manuscript.

BS conducted the laboratory observations and coding, contributed to the data analysis, and the write up of the manuscript.

HM contributed to the data analysis and the write up of the manuscript.

## Funding

This study was funded by the Research Partnership for Holistic Measurement of Education Results, New York University.

## Availability of data and materials

Raw data and materials are available under request at the office of the Principal Investigator: Leyla Akoury Dirani, [la55@aub.edu.lb](mailto:la55@aub.edu.lb).

## Declarations

### Ethics approval and consent to participate

The research study was granted the approval of the Institutional Review Board of the American University of Beirut. All participants signed informed consent forms.

### Consent for publication

All authors consented to this publication.

### Competing interests

All authors declare that they have no competing interests.

Received: 15 April 2024 Accepted: 8 May 2024

Published online: 14 June 2024

## References

1. Anzman-Frasca S, Stifter CA, Paul IM, Birch LL (2014) Negative temperament as a moderator of intervention effects in infancy: testing a differential susceptibility model. *Prev Sci* 15(5):643–653. <https://doi.org/10.1007/s11121-013-0408-4>
2. Brooker RJ, Kiel EJ, Buss KA (2016) Early social fear predicts kindergarten’s socially anxious behaviors: Direct associations, moderation by inhibitory control, and differences from nonsocial fear. *Emotion* 16(7):997–1010. <https://doi.org/10.1037/emo0000135>
3. Buss KA (2011) Which fearful toddlers should we worry about? Context, fear regulation, and anxiety risk. *Dev Psychol* 47(3):804–819. <https://doi.org/10.1037/a0023227>
4. Buss K.A, Goldsmith H.H (2000) Manual and normative data for the Laboratory Temperament Assessment Battery – Toddler Version. In. Psychology Department Technical Report, University of Wisconsin–Madison.
5. Capaldi DM, Rothbart MK (1992) Development and validation of an early adolescent temperament measure. *J Early Adolesc* 12:153–173
6. Cicchetti DV (1994) Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess* 6(4):284–290. <https://doi.org/10.1037/1040-3590.6.4.284>
7. Cosentino-Rocha L, Linhares M (2013) Child temperament and gender differences Temperamento de Crianças e Diferenças de Gênero. *Paidéia (Ribeirão Preto)* 23:63–72. <https://doi.org/10.1590/1982-43272354201308>
8. Danzig AP, Dyson MW, Olino TM, Lappook RS, Klein DN (2015) Positive parenting interacts with child temperament and negative parenting to predict children’s socially appropriate behavior. *J Soc Clin Psychol* 34(5):411–435. <https://doi.org/10.1521/jscp.2015.34.5.411>
9. De Pauw SS, Mervielde I (2010) Temperament, personality and developmental psychopathology: a review based on the conceptual dimensions underlying childhood traits. *Child Psychiatry Hum Dev* 41(3):313–329. <https://doi.org/10.1007/s10578-009-0171-8>
10. DeVries MW, Sameroff AJ (1984) Culture and temperament: influences on infant temperament in three East African societies. *Am J Orthopsychiatry* 54(1):83–96. <https://doi.org/10.1111/j.1939-0025.1984.tb01477.x>
11. Evans DE, Rothbart M (2007) Developing a model for adult temperament. *J Res Pers* 41:868–888

12. Faisca L, Ferreira LI, Fernandes CC, Gagne JR, Martins AT (2021) Behavioral inhibition in childhood: European Portuguese Adaptation of an Observational Measure (Lab-TAB). *Children (Basel)* 8(2):162
13. Gagne JR, Miller MM, Goldsmith HH (2013) Early-but modest-gender differences in focal aspects of childhood temperament. *Personality Individ Differ* 55(2):95–100. <https://doi.org/10.1016/j.paid.2013.02.006>
14. Gaias LM, Rääkkönen K, Komsí N, Gartstein MA, Fisher PA, Putnam SP (2012) Cross-cultural temperamental differences in infants, children, and adults in the United States of America and Finland. *Scand J Psychol* 53(2):119–128. <https://doi.org/10.1111/j.1467-9450.2012.00937.x>
15. Gandour M (1989) Activity level as a dimension of temperament in toddlers: its relevance for the organismic specificity hypothesis. *Child Dev* 60(5):1092–1098. <https://doi.org/10.2307/1130783>
16. Gartstein M, Rothbart M (2003) Studying infant temperament via the Revised Infant Behavior Questionnaire. *Infant Behav Dev* 26:64–86
17. Goldsmith H.H, Lemery K.S, Schmidt N.L, Chow O.K, Justen A.A (2010) Field administration manual for the Laboratory Temperament Assessment Battery: Middle Childhood Version. In. Psychology Department Technical Report, University of Wisconsin–Madison
18. Goldsmith H.H, Reilly J, Lemery K.S, Longley S, Prescott A (1993) Preliminary manual for the preschool laboratory temperament assessment battery (version 1.0). In. Technical Report, Department of Psychology, University of Wisconsin–Madison.
19. Goldsmith HH, Rothbart MK (1991) Contemporary instruments for assessing early temperament by questionnaire and in the laboratory. In: Strelau J, Angleitner A (eds) *Explorations in temperament: International perspectives on theory and measurement*. Springer, New York, NY, pp 249–272. [https://doi.org/10.1007/978-1-4899-0643-4\\_16](https://doi.org/10.1007/978-1-4899-0643-4_16)
20. Goldsmith HH, Rothbart MK (1996a) The Laboratory Temperament Assessment Battery (LAB-TAB): Locomotor Version 3.0. Technical manual, Department of Psychology, University of Wisconsin, Madison, WI.
21. Goldsmith HH, Rothbart MK (1996b) The Laboratory Temperament Assessment Battery (LAB-TAB): Prelocomotor Version 3.0. In. Technical manual, Department of Psychology, University of Wisconsin, Madison, WI.
22. Goldsmith HH, Rothbart MK (1999) The Laboratory Temperament Assessment Battery (LAB-TAB): Locomotor Version 3.1. Madison, WI: University of Wisconsin.
23. Grant S (2008) Validation of an Ethiopian Version of the Laboratory Temperament Assessment Battery Southern Nazarene University
24. Hane AA, Fox NA, Polak-Toste C, Ghera MM, Guner BM (2006) Contextual basis of maternal perceptions of infant temperament. *Dev Psychol* 42(6):1077–1088. <https://doi.org/10.1037/0012-1649.42.6.1077>
25. Hsin A, Felfe C (2014) When does time matter? maternal employment, children's time with parents, and child development. *Demography* 51(5):1867–1894. <https://doi.org/10.1007/s13524-014-0334-5>
26. Ishii K, Shibata A, Adachi M, Nonoue K, Oka K (2015) Gender and grade differences in objectively measured physical activity and sedentary behavior patterns among Japanese children and adolescents: a cross-sectional study. *BMC Public Health* 18(15):1254. <https://doi.org/10.1186/s12889-015-2607-3.PMID:26679503;PMCID:PMC4683705>
27. Jansen PW, Raat H, Mackenbach JP, Jaddoe VW, Hofman A, Verhulst FC, Tiemeier H (2009) Socioeconomic inequalities in infant temperament: the generation R study. *Soc Psychiatry Psychiatr Epidemiol* 44(2):87–95. <https://doi.org/10.1007/s00127-008-0416-z>
28. McGraw KO, Wong SP (1996) Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1(1):30–46. <https://doi.org/10.1037/1082-989X.1.1.30>
29. Planalp EM, Van Hulle C, Gagne JR, Goldsmith HH (2017) The infant version of the laboratory temperament assessment battery (Lab-TAB): Measurement properties and implications for concepts of temperament. *Front Psychol* 8(846):1–17. <https://doi.org/10.3389/fpsyg.2017.00846>
30. Putnam SP, Gartstein MA, Rothbart MK (2006) Measurement of fine-grained aspects of toddler temperament: the early childhood behavior questionnaire. *Infant Behav Dev* 29(3):386–401. <https://doi.org/10.1016/j.infbeh.2006.01.004>
31. Putnam SP, Jacobs J, Gartstein MA, Rothbart MK (2010) Development and assessment of short and very short forms of the Early Childhood Behavior Questionnaire. Poster presented at International Conference on Infant Studies, Baltimore, MD.
32. Rothbart MK (1980) Longitudinal home observation of infant temperament International Conference on Infant Studies, New Haven, CT.
33. Rothbart MK, Ahadi SA, Hershey KL, Fisher P (2001) Investigations of temperament at three to seven years: the children's behavior questionnaire. *Child Dev* 72(5):1394–1408. <https://doi.org/10.1111/1467-8624.00355>
34. Saudino KJ (2005) Behavioral genetics and child temperament. *J Dev Behav Pediatr* 26(3):214–223. <https://doi.org/10.1097/00004703-200506000-00010>

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.