Published in final edited form as:

J Pers Assess. 2011 November ; 93(6): 545–555. doi:10.1080/00223891.2011.608756.

Psychometric Properties of the Behavioral Inhibition Questionnaire in Preschool Children

Jiyon Kim¹, Daniel N. Klein¹, Thomas M. Olino², Margaret W. Dyson¹, Lea R. Dougherty³, and C. Emily Durbin⁴

¹Department of Psychology, Stony Brook University

²Department of Psychiatry, University of Pittsburgh School of Medicine

³Department of Psychology, University of Maryland, College Park

⁴Department of Psychology, Northwestern University

Abstract

We examined the psychometric properties of the Behavioral Inhibition Questionnaire (BIQ), a rating scale for children's behavioral inhibition (BI). Parent and teacher ratings, parent interviews, and laboratory observations were obtained for 495 preschoolers. Confirmatory factor analysis yielded six factors, each reflecting the BIQ's subscales, and all loading onto a second-order general dimension. Model fit was acceptable for parent ratings, but only marginal for teacher ratings. The convergent and discriminant validity of the BIQ was examined by using a multitrait multimethod approach. Results indicate that the BIQ displays evidence of reliability and validity that can complement observational paradigms.

Behavioral inhibition (BI) refers to a temperamental tendency to display wariness, fearfulness, or restraint in response to unfamiliar people, objects, and situations (Kagan, Reznick, Snidman, Gibbons, & Johnson, 1988). When exposed to the unfamiliar, inhibited children may respond by seeking proximity to their caregivers and show reduced approach towards novel stimuli or people. The construct of BI differs from most other current models of temperament in several ways (Zentner & Bates, 2008). First, while most major theories of temperament adopt a multidimensional framework, Kagan and colleagues (e.g., Kagan, Reznick, and Snidman, 1988) have focused on a single trait. Second, most current approaches to child temperament originated from theoretical models of temperament. In contrast, Kagan's work on BI is inductive, developing from a series of empirical studies on the continuity and stability of behavior beginning in early childhood (Kagan & Snidman, 2004). For example, they found that many 2-year-olds categorized as inhibited based on reactivity to unfamiliar situations continued to display inhibited behavior into adolescence (Kagan & Snidman, 2004). While current models of temperament include a variety of constructs, three traits are common to almost all of them: surgency/positive emotionality (PE), negative emotionality (NE), and constraint/effortful control (Caspi & Shiner, 2006; De Pauw & Mervielde, 2010; Rothbart & Bates, 2006). BI overlaps with, but also differs in important ways from each of these three traits. BI and low surgency/PE share a low level of approach/engagement. However, the approach deficit in the latter is assumed to be consistent across situations, whereas in BI it is limited to novel situations. (Laptook, Klein, Olino, Dyson, & Carlson, 2010; Polak-Toste & Gunnar, 2006). Moreover, in the social/interpersonal domain, low surgency/PE may be manifested by a lack of desire to interact

with others, whereas behaviorally inhibited, or shy, children often desire social interaction but are inhibited by wariness and anxiety (Coplan & Armer, 2007). BI and NE are both characterized by fearfulness and anxiety, but NE is also characterized by other dysphoric affects, such as irritability and sadness. In addition, in BI, fear is associated with novel contexts, whereas NE is conceptualized as a generalized trait, with less attention paid to eliciting cues and contexts. Similarly, while BI and high constraint/effortful control are both characterized by caution and restraint, the latter is more generalized, while the former is evident only in unfamiliar contexts (Durbin, Klein, Hayden, Buckley, & Moerk, 2005).

The construct of BI also differs from approach-withdrawal theories of temperament, despite some conceptual overlap. Kagan's notion of BI was developed independently from biobehavioral models of motivation and emotional reactivity (e.g. Davidson, 1992; Gray, 1991). For example, Gray's (1991) model of the Behavioral Inhibition System (BIS) and Behavioral Approach System (BAS) proposes distinct neurobiological systems that underlie motivational behavior. The BAS regulates approach behaviors in response to cues for reward, whereas the BIS activates withdrawal behaviors in response to cues for punishment, novel stimuli, and non-reward. BI overlaps conceptually with Gray's BIS. However, unlike the BIS, signals of non-reward should not elicit BI. Moreover, Kagan's theory is more explicitly developmental and focused on individual differences (Zentner & Bates, 2008).

A series of studies has shown that BI can be observed as early as infancy (Kagan & Snidman, 2004), and appears to have moderate stability from infancy to early childhood (Broberg, Lamb, & Hwang, 1990; Fox, Henderson, Rubin, Calkins, & Schmidt, 2001), early to late childhood (Asendorpf, 1994; Scarpa, Raine, & Venables, 1995), and, to a somewhat smaller extent, from infancy into adolescence (Kagan, Snidman, Kahn, & Towsley, 2007). BI also appears to have a number of neurobiological correlates. For example, behaviorally inhibited infants and preschoolers exhibit greater right than left frontal cortical activity on the electroencephalogram (Calkins, Fox, & Marshall, 1996; Fox et al, 2001), which is also evident in late childhood-early adolescence (McManis, Kagan, Snidman, & Woodward, 2002). In addition, infants and toddlers with BI exhibit heightened startle response (Schmidt & Fox, 1998), and as adolescents they show increased amygdala activation in response to fear stimuli (Perez-Edgar et al., 2007; Schwartz, Wright, Shin, Kagan, & Rauch, 2003). Finally, BI in early childhood appears to be a risk factor for the development of anxiety disorders in middle and late childhood and adolescence (Chronis-Toscano et al., 2009; Hirshfeld-Becker et al., 2008). These findings suggest that BI in early childhood may be a good target for prevention and early intervention programs aimed at anxiety disorders (Rapee, Kennedy, Ingram, Edwards, & Sweeney, 2005).

BI in children is typically assessed in the laboratory by exposing children to a series of novel objects or unfamiliar adults or peers. Indicators of inhibition include a longer latency to approach unfamiliar people or objects, decreased frequency and duration of interaction or play, more referencing and maintaining closer proximity to the caregiver, and fearful or avoidant behavior. However, observational methods can be time-consuming and expensive. Furthermore, no single observational procedure has been accepted as the gold standard and protocols vary widely across different laboratories (Rothbart & Bates, 2006).

While structured observations are the standard method of assessing BI, some researchers have used concurrent reports by parents, teachers, and older children, and retrospective reports by adolescents and adults. These measures have several strengths that complement laboratory observations. Questionnaires can capture behaviors across longer intervals than those displayed within a limited observational period. In addition, use of different informants (e.g., parents, teachers) allows for observations across multiple contexts. Further,

questionnaires are economical and efficient, and therefore are well suited for screening and as secondary measures in larger studies. However, questionnaires also have significant limitations. They are prone to various biases stemming from differences in frame of reference, interpretation and understanding of questions, and social desirability. Thus, some researchers have strongly argued against the use of parent report in assessing BI (Kagan, 1998). In light of the differing strengths and weaknesses of observational and questionnaire measures, many researchers advocate using both approaches to provide a more comprehensive picture of child temperament (Mangelsdorf, Schoppe, & Buur, 2000; Rothbart & Bates, 2006).

Unfortunately, there are only a few questionnaire measures designed specifically to assess BI in children, and validation data on these measures are limited. The two best-documented measures are the Behavioral Inhibition Scale (BIS; Muris, Meesters, & Spinder, 2003) and the Behavioral Inhibition Questionnaire (BIQ; Bishop, Spence, & McDonald, 2003). The BIS has a parent version and a self report version that can be used for older children. However, all BIS items assess behavior in social contexts; inhibited behaviors in non-social situations are not assessed.

The BIQ is a more comprehensive parent and teacher rating scale developed for use with preschoolers. The BIQ assesses children's inhibited behavior in a greater range of contexts than the BIS, including with unfamiliar adults, with unfamiliar peers, in unfamiliar situations, in performance situations, in separation/preschool situations, and with novel physical activities that may entail minor risks. Thus, the BIQ's six subscales provide good coverage of different aspects of the construct.

Using a large Australian sample, Bishop et al. (2003) reported that the BIQ had good psychometric properties. Confirmatory factor analyses of mother, father and teacher reports supported a structure consisting of six correlated factors, each corresponding to one of the six contexts assessed by the BIQ's subscales, all of which also loaded onto a single higher-order construct. Both the parent and teacher versions were found to have acceptable internal consistency, with alphas for the total and factor scores above .80, except for father report on the physical challenge factor ($\alpha = .72$). Parent reports had satisfactory test-retest reliability (r s ranging from .60–.78 for mothers and .55–.74 for fathers). The agreement between different informants was relatively good, with greater convergence between mother and father ($r = .69$) than with teacher report ($r = .47$ and .43 for mother and father). Further, concurrent and construct validity of the BIQ were found to be satisfactory. The BIQ total scores were strongly associated with the inhibition subscale of the Temperament Assessment Battery for Children – Revised (TABC – R; Presley & Martin, 1994), with r s ranging from .85–.87 (all $ps < .001$). Significant correlations were also found between BIQ ratings from all three informants and observational ratings of BI in a laboratory setting (r s .25–.46 for full sample and .35–.60 for high/low BI groups).

The goal of the present study is to further examine the structure and psychometric properties of the parent and teacher versions of the BIQ in a large sample of preschool-aged children in the United States. We aim to replicate Bishop et al.'s (2003) findings, and extend them by using a broader range of convergent measures of BI and related constructs and assessing the BIQ's discriminant validity. First, we assessed the internal consistency of the BIQ. Second, we attempted to replicate the factor structure obtained in the original study. Next, the convergent and discriminant validity of the BIQ were examined using a multitrait-multimethod (MTMM) approach (Campbell & Fiske, 1959). We included traits that were closely related to BI (e.g. fear, shyness) and traits related to low surgency/PE (low positive affect, social disinterest), high NE (e.g., negative emotional reactivity, anger/frustration, sadness), and high constraint (e.g., low impulsivity), that should exhibit some overlap with,

but still be relatively distinct from, BI. These analyses included comparisons with other questionnaire measures, a standardized observational protocol, global experimenter ratings, and a semi-structured parent interview.

Methods

Participants

Participants were 559 parents and their three-year-old children from Long Island, NY. The mean age of the children was 3.5 years ($SD = 0.3$). The majority of the children in the study were Caucasian (87%) and slightly less than half were female (46%). Parents were primarily married (94%) and 51% of mothers worked outside the home. Fifty-five percent of the mothers and forty-seven percent of the fathers had a four-year college or graduate degree. Participants were recruited through a commercial mailing list. Families with a three-year-old child who lived with at least one English-speaking biological parent and did not have any significant medical conditions or developmental disabilities were eligible to participate. Of the total sample, 495 parents (88.6%) complete the BIQ parent version. For children who were enrolled in preschool or daycare (71% of the sample, $N = 397$), teachers were asked to complete questionnaires. Of those, 226 teachers (56.9%) completed the BIQ. To examine selection bias, children with and without teacher reports were compared on parent reports on the BIQ. The groups did not differ on the total BIQ scale or any of the six subscales.

Measures

Behavioral Inhibition Questionnaire (BIQ)—The BIQ (Bishop, Spence, & McDonald, 2003) assesses the frequency of children's behavioral inhibition across six contexts in three domains: social novelty (unfamiliar adults, peers, and performing in front of others), situational novelty (unfamiliar situations, preschool/separation), and novel physical activities with possible risk of injury. The authors first aggregated a list of 40 items based on a review of existing literature on BI across different contexts and existing temperament questionnaires. Items with low understandability and poor psychometric properties were discarded. The remaining 30 items were given to two experienced clinical child psychologists with expertise in child BI. They were asked to allocate the items into one of the three domains (social, situational, and physical risk). Their allocation of items closely resembled that hypothesized by the authors. The draft questionnaire was administered to a pilot group of parents and teachers for feedback on readability and understandability, after which minor changes were made to produce the final versions. Items are rated on a seven-point scale ranging from 1 (*hardly ever*) to 7 (*almost always*). Item ratings are summed to create six scale scores – unfamiliar peers, unfamiliar adults, performance situations, separation/preschool, unfamiliar situations, and physical challenge situations. A parent form, consisting of 30 items, and a teacher form, consisting of 28 items (excluding two items on children's behaviors in unfamiliar homes), were completed by a parent (474 mothers and 21 fathers) and teacher ($N = 226$), respectively.

Children's Behavior Questionnaire (CBQ)—The CBQ (Rothbart, Ahadi, Hersey, & Fisher, 2001) is a widely used 195-item caregiver report measure of temperament for three- to seven-year-old children. Parents are asked to rate the child's reaction/behavior within the past six months on a seven-point scale ranging from 1 (*extremely untrue of your child*) to 7 (*extremely true of your child*). 474 mothers and 21 fathers completed the CBQ. The CBQ provides dimensional scores on 16 scales based on the mean ratings of pertinent items. Scales used in the present analyses included fear (12 items, $\alpha = .74$) and shyness (13 items, $\alpha = .92$). In addition, the anger/frustration (13 items, $\alpha = .79$) and smiling/laughter (13 items, $\alpha = .73$) scales were included to examine the discriminant validity of the BIQ scales.

Child Social Preference Scale (CSPS)—The CSPS (Coplan, Prakash, O'Neill, & Armer, 2004) consists of 11 items designed to assess children's conflicted shyness and social disinterest. Both a parent (471 mothers and 21 fathers) and teacher ($N = 205$) completed this measure by indicating the degree to which each item reflects the child's general social behavior. Items are rated on a five-point scale ranging from 1 (*not at all*) to 5 (*a lot*). Scale scores are computed by averaging the items. In this study, coefficient alphas for the parent-rated shyness (N of items = 7) and disinterest scales (N of items = 4) were .90 and .73, respectively. For teachers, alphas were .88 and .86, respectively.

Children's Reaction Scale (CRS)—The CRS (Eisenberg et al., 1996) is a 19-item measure that assesses positive, negative, and general emotionality in children. Each item is rated on a seven-point scale ranging from 1 (never) to 7 (always) and summed to create scale scores. Only the items ($N = 5$) pertaining to negative emotionality were used in this study. The CRS was completed by both a parent and the teacher. The coefficient alphas for the parent and the teacher forms were .70 and .75, respectively.

Laboratory Temperament Assessment—During their visit to the laboratory, children participated in a standardized set of twelve episodes from the Laboratory Temperament Assessment Battery (Lab-TAB; Goldsmith, Reilly, Lemery, Longley, & Prescott, 1995), lasting approximately two hours. Three of these episodes were designed to assess BI (*Risk Room*, *Stranger Approach*, and *Exploring New Objects*), and are similar to those used in previous studies of this trait (e.g., Kochanska & Knaack, 2003; Pfeifer et al., 2002). Each episode was videotaped through a one-way mirror and later coded. A parent remained in the room except for Stranger Approach.

1. *Risk Room*. The child explored a set of novel and ambiguous stimuli (e.g., cloth tunnel, balance beam, Halloween mask, etc.).
2. *Tower of Patience*. The child and the experimenter took turns building a tower of large cardboard blocks. The experimenter gradually delayed placing her block, forcing the child to wait longer each time for his or her turn.
3. *Arc of Toys*. The child played freely by him- or herself in a roomful of toys for a few minutes, after which the experimenter returned and instructed the child to clean up the toys.
4. *Stranger Approach*. While the experimenter went to get more toys, a male research assistant entered the room where the child waited alone and spoke to the child in a neutral tone while gradually walking closer.
5. *Car Go*. The child and the experimenter played with two remote-controlled racecars.
6. *Transparent Box*. The experimenter locked an appealing toy of the child's choice in a transparent box and left the child with a set of inoperable keys to open the box. After a few minutes, the experimenter returned and explained that she had accidentally given the wrong keys. The experimenter then encouraged the child to play with the toy.
7. *Exploring New Objects*. The child explored a set of novel and ambiguous stimuli, including a mechanical spider, a mechanical bird, and sticky water-filled soft gel balls.
8. *Pop-up Snakes*. The child and the experimenter surprised the child's mother or father using a coiled spring snakes hidden in what appears to be a can of potato chips.

9. *Impossibly Perfect Green Circles*. The child was repeatedly asked to draw a circle on a large piece of paper, while each circle was mildly criticized by the experimenter.
10. *Popping Bubbles*. The child and the experimenter played together blowing bubbles.
11. *Snack Delay*. The child was instructed to wait for the experimenter to ring a bell before getting a snack. The experimenter took increased delays before ringing the bell.
12. *Box Empty*. The child was left alone to open a wrapped gift box that turned out to be empty. The experimenter returned after a few minutes, explained her mistake, and gave several toys for the child to keep.

Tape Coding Procedures—BI was coded using an approach that is similar to most previous studies (e.g., Kagan et al., 1988; Pfeifer, Goldsmith, Davidson, & Rickman, 2002). The three episodes specifically designed to assess BI were divided into 20 or 30 second epochs, and a series of affective and behavioral codes were rated for each epoch (Goldsmith et al., 1995). Within each epoch, a maximum intensity rating of facial, bodily, and vocal fear was coded on a scale of 0 (absent) to 3 (highly present and salient). Facial affect and their intensity were coded using Izard et al.'s (1989) AFFEX system. Based on previous studies using the Lab-TAB (Durbin et al., 2005; Pfeiffer et al., 2002), BI was computed as the average standardized ratings of latency to fear (reversed); and facial, vocal, and bodily fear (*Risk Room*, *Stranger Approach*, and *Exploring New Objects*); latency to touch objects; total number of objects touched (reversed); tentative play; referencing the parent; proximity to parent; referencing the experimenter; and time spent playing (reversed) (*Risk Room* and *Exploring New Objects*); startle (*Exploring New Objects*); sad facial affect (*Exploring New Objects* and *Stranger Approach*); and latency to vocalize; approach towards the stranger (reversed); avoidance of the stranger; gaze aversion; and verbal/nonverbal interaction with the stranger (reversed; *Stranger Approach*).

Positive affect, sadness, anger, and impulsivity were coded in each of the 12 episodes based on a system developed by Durbin, Hayden, Klein, and Olino (2007) for the Lab-TAB. Each display of facial, bodily and vocal positive affect, sadness, and anger in each episode was rated on a three-point scale (low, moderate, high). Ratings were summed separately within each channel (facial, bodily, vocal) across the 12 episodes, standardized, and summed across the three channels to derive total scores for positive affect, sadness, and anger. These scores can be as low as 0 and do not have upper limits. Impulsivity was rated on a single four-point scale (none, low, moderate, and high) for each episode based on the quantity and quality of the child's tendency to act or respond without reflection or hesitation. These ratings were then averaged across the 12 episodes. Impulsivity ratings can range from 0 to 4.

For each episode, the raters who coded BI differed from the raters who coded positive affect, sadness, anger, and impulsivity. BI exhibited good internal consistency ($\alpha=.80$) and inter-rater reliability ($ICC=.88$; $N=28$). Internal consistencies for positive affect, sadness, anger, and impulsivity were .87, .81, .68, and .70, respectively. Interrater reliabilities ($N = 28$) were .92 for positive affect, .79 for sadness, .73 for anger, and .75 for impulsivity.

Experimenter ratings—Immediately after administering the Lab-TAB, the experimenter completed 24 items rating the child's behavior throughout the observation session. These global ratings were based on the experimenter's interaction with the child throughout the whole visit, including the periods before, during, between, and after the 12 episodes. For the present study, a total BI rating was constructed by computing the mean of three items on shyness, fear (objects/situations), and exploration of objects (reverse coded). Coefficient

alpha for this scale was .62. We were not able to assess interrater reliability for the experimenter ratings.

Preschool Age Psychiatric Assessment (PAPA)—The PAPA (Egger, Ascher, & Angold, 1999) is a semi-structured interview assessing parent-reported psychopathology in preschool children between the ages of 2 and 5 years. The PAPA covers a comprehensive set of symptoms from the *Diagnostic and Statistical Manual of Mental Disorders, 4th edition, Text Revision* (DSM-IV-TR; American Psychiatric Association, 2000) as well as developmentally relevant items, such as eating, sleeping, and play behaviors, that are absent from the current nosology. Adequate test-retest reliability has been found for the PAPA (Egger et al., 2006). The PAPA includes five items assessing BI (global impression of BI, BI with unfamiliar adults, BI with unfamiliar children, BI in new situations, and BI with new objects). These items were summed to provide a BI scale ($\alpha = .85$; interrater ICC = .99, $N = 21$). Parents were interviewed (482 mothers, 7 fathers, and 1 both parents) by graduate students in clinical psychology who received training on the administration of the PAPA by its developers. Interviews were conducted by telephone.

Results

Internal Consistency

Table 1 presents the internal consistency of the total score and the six factor scores. Good internal consistency ($> .80$) was found for all scales reported by parents and teachers, except for the parent-reported physical challenges factor (.74).

Relationship between Parent and Teacher Reports of BIQ

The BIQ total scores rated by parents and teachers were significantly correlated ($r = .49$, $p < .001$, $n = 206$). For the subscales, significant correlations were found for unfamiliar peers ($r = .38$, $p < .001$), unfamiliar adults ($r = .42$, $p < .001$), performance situations ($r = .46$, $p < .001$), separation/preschool ($r = .43$, $p < .001$), and unfamiliar situations ($r = .39$, $p < .001$). The correlation for physical challenge situations was not statistically significant ($r = .09$, $p = .20$). Table 2 presents intercorrelations between parent and teacher ratings across all subscales and total scores. For those children with both parent and teacher reports ($n = 210$), we compared the mean ratings on five subscales and the total score. For this analysis, parent reports were rescored without the two items in unfamiliar situations that are not included in the teacher version. Results from paired t -tests show that the mean teacher ratings were significantly higher than the mean parent ratings on unfamiliar peers ($t = 2.39$, $p = .02$, $df = 209$, Cohen's $d = .18$) and performance situations ($t = 3.81$, $p < .001$, $df = 209$, Cohen's $d = .27$). The remaining subscales showed no statistically significant differences between teacher and parent ratings (Cohen's d all $< .11$).

Gender Differences

Table 1 reports the means, standard deviations, and effect sizes for gender differences for total and six subscale scores. Boys and girls did not differ on BIQ total scores according to either parents or teachers. However, there were some gender differences on the BIQ subscales. Boys were rated by their parents as more inhibited in performance situations, $F(1, 493) = 9.63$, $p < .01$, and preschool/separation than girls, $F(1, 493) = 9.64$, $p < .01$. Girls were rated by their parents as significantly more inhibited in novel physical activities, $F(1, 493) = 8.86$, $p < .01$. Teachers rated boys as significantly more inhibited in performance situations, $F(1, 225) = 6.23$, $p < .01$.

Confirmatory Factor Analysis

To examine the factor structure of the BIQ, confirmatory factor analysis with maximum-likelihood estimation procedures was conducted for parent and teacher reports. Analyses were conducted using the Amos 7.0 software package (Arbuckle, 2006). To reduce skewness and kurtosis, we deleted four cases that were extreme multivariate outliers and transformed 12 variables using square root and log transformation for moderate and substantial skewness, respectively (Watches Children, Many Days to Adjust, Hesitant in New Situations, Settles In Quickly, Independent, Comfortable in New Situations, Clingy, Happily Approaches New Activities, Outgoing, Nervous in New Situations, Explores Play Equipment, Hesitant with New Play Equipment).¹ In order to replicate the original validation study by Bishop and colleagues (2003), we followed the same analytic procedures and compared four models for parent and teacher reports: (a) single factor, (b) three correlated factors, (c) six correlated factors, and (d) six correlated factors loading onto one higher order factor. The first model proposed that all questionnaire items across all dimensions load onto a single dimension. The second model hypothesized that the items cluster around three contexts in which inhibited behavior occurs (social novelty, situational novelty, and novel physical activities with possible risk of injury). The third model posited that the items cluster around the six domains that comprise the three contexts (peer situations, unfamiliar adults, performing in front of others, unfamiliar situations, preschool/separation, and physical challenge). The final model proposed that the covariance among the six first-order factors can be explained by a single second-order factor of overall BI. For all analyses, we analyzed the covariance matrix. Each of the latent variable metrics was set by fixing one factor loading at 1 and we did not specify any residual correlations between the indicators. Model fit was evaluated by using multiple fit indices, including the chi-square statistic, the Comparative Fit Index (CFI), and the Root Mean Square Error of Approximation (RMSEA). Chi-square evaluates the fit between the covariance matrix of the sample and the estimated population, and a nonsignificant value indicates good fit of the model. However, it is sensitive to sample size, so that with large samples trivial discrepancies often lead to rejection of an otherwise satisfactory model. Therefore, we gave greater weight to the CFI and the RMSEA, which are the most frequently reported fit indices and among the measures least affected by sample size. The CFI compares the proposed model fit with a null model that assumes the latent variables are uncorrelated. CFI values greater than .95 indicate a good fit (Hu & Bentler, 1999) whereas .90 suggests an adequate fit. The RMSEA estimates the lack of fit in a model compared to a perfect model, relative to the model degrees of freedom. Values less than .05 indicate a very good fit, and values below .08 indicate an adequate fit (Byrne, 2001; Loehlin, 2004).

Model 1 (single factor)—All items loaded significantly for both parent and teacher reports, with item loadings above .40 for all teacher report items and 27 of the 30 parent report items. As can be seen in Table 3, however, the fit indices suggest that this model does not represent a good fit to the data.

Model 2 (3 correlated factors)—All items loaded greater than .50 on their hypothesized factors for both parent and teacher reports. Compared to model 1, the three-factor model (social novelty, situational novelty, and novel physical activities with possible risk of injury) provided a better fit; however, the fit indices indicate that this solution is also a poor fit.

Model 3 (6 correlated factors)—In comparison to model 2, the six correlated factor model provided a significantly better fit to the data for both parent and teacher reports. While the Chi-square value was significant, the CFI and RMSEA indicated that the parent

¹The results of all other analyses were virtually unchanged when the four outliers were omitted.

model provided a relatively good fit to the data, whereas the teacher model provided a marginal fit. The six latent factors were all correlated with one another ($p < .001$). All item loadings were statistically significant in the predicted direction and relatively large in magnitude for both parent and teacher reports. These results provide support for a model in which BI responses cluster around the six contexts (approaching unfamiliar peers, approaching unfamiliar adults, performing in front of others, separation and preschool situations, unfamiliar situations in general, and novel physical activities suggestive of minor risks) in both parent and teacher reports.

Model 4 (6 first-order factors, 1 second-order factor)—As shown in Table 3, the second-order factor model also fits the data well for parent reports, indicating that the six context factors load onto a single factor of overall BI.² However, the teacher model provided only a marginal fit again. In order to compare the second-order model to the first-order model, we used Marsh and Hocevar's (1985) target coefficient, which is the ratio of the chi-square of the first-order model to that of the second-order model. It has an upper limit of 1, which indicates that the covariation among the first-order factors can be completely explained by the second-order factor. A target coefficient above .90 suggests that the second-order factor accounts for a large portion of the covariance among the first-order factors. In the present dataset, the target coefficient between the six-factor model and the second-order model was .97 for parent reports and .95 for teacher reports, indicating that the higher order model accounts for almost all of the covariance among the six first-order factors. For both parent and teacher reports, all item loadings were significant and above .50. The loadings of all six first-order factors onto the second-order factors were also significant (see Figure 1).

Convergent and Discriminant Validity of the BIQ

A multitrait-multimethod approach was used to examine convergent and discriminant validity of the BIQ. Analyses were performed separately for the parent- and teacher-reported BIQ. As it can be debated whether measures using the same method but different informants should be considered monomethod or heteromethod comparisons, we did not examine these correlations (i.e., teacher questionnaire measures were not included in the convergent and discriminant analyses for the parent BIQ and parent questionnaire measures were not included in the analyses for the teacher BIQ).

Parent Reports—Table 4 displays the convergent and discriminant associations for the parent BIQ. Convergent measures included parent questionnaire indices of shyness from the CBQ and the CSPS and fear from the CBQ, and measures of BI assessed using a semi-structured interview with a parent (PAPA), behavior-based codes from a structured laboratory observation protocol (Lab-TAB), and experimenter global ratings from the same laboratory visit. For discriminant validity, traits related to, but distinct from BI, were examined. These include traits reflecting high NE, low surgency/PE, and high constraint assessed with parent questionnaires (CBQ anger/frustration and smiling/laughter; CRS negative emotionality, and CSPS social disinterest), and a structured laboratory observation (Lab-TAB sadness, anger, positive affect, and impulsivity).

With only a few exceptions, the pattern of correlations supported the convergent and discriminant validity of the BIQ: monotrait-monomethod correlations with the CBQ and the

²For parent report, the error variance associated with the Unfamiliar Situations factor was negative, but not statistically significantly different from zero. We fixed this parameter to zero after careful examination of the dataset and the model ruled out any reason to suspect model misspecification. Monte Carlo studies have shown that non-significant negative error variance estimates should not be regarded as an indicator of model misspecification, and that fixing such estimates to zero may reduce bias in parameter estimates (Chen, Bollen, Paxton, Curran, & Kirby, 2001).

CSPS were generally greater than monotrait-heteromethod correlations among different methods of assessing BI, most of which were larger than the heterotrait-monomethod correlations, which, in turn, were generally greater than the heterotrait-heteromethod correlations.

All monotrait-monomethod correlations were statistically significant. Correlations with the shyness subscales from the CBQ and the CSPS were substantial (.85 and .78, respectively), whereas the correlation with the fear subscale of the CBQ was considerably lower (.30). Monotrait-heteromethod correlations with BIQ scores were low to moderate, ranging from .25 to .63. Not surprisingly, PAPA BI was more highly correlated with parent BIQ scores than observed BI ratings, probably because with only a few exceptions, the same informant completed the PAPA and parent BIQ. BIQ scores were more highly correlated with experimenters' global ratings than with the Lab-TAB behavioral scores. This may reflect the fact that the experimenter ratings were based on the entire two-hour laboratory visit (including the child's arrival prior to beginning the protocol), whereas the behavioral scores were based on only three of the 12 episodes in the laboratory session. Heterotrait-monomethod correlations varied across traits, ranging from low to moderate (.16 to $-.41$). Finally, heterotrait-heteromethod correlations were uniformly low ($-.05$ to $-.18$), and smaller than all of the monotrait-monomethod and monotrait-heteromethod correlations.

Teacher Reports—Convergent measures included teacher reported shyness from the CSPS and BI assessed with a semi-structured interview with a parent (PAPA), a laboratory protocol (Lab-TAB), and experimenter global ratings. Discriminant traits sharing the same method were teacher reports of social disinterest on the CSPS and negative emotionality on the CRS. Discriminant traits using different method were assessed through a structured laboratory observation (Lab-TAB sadness, anger, positive affect, and impulsivity).

As can be seen in Table 5, the monotrait-monomethod correlation with the shyness subscale of the CSPS was substantial and significant. Monotrait-heteromethod correlations with the teacher BIQ were in the low to moderate range (.15–.46). Similar to the parent report, teacher-reported BIQ showed the strongest correlation with the PAPA BI (.46), followed by experimenter-rated BI and Lab-TAB BI. The two heterotrait-monomethod correlations were low (CRS negative emotionality) to moderate (CSPS social disinterest). The latter correlation (.45) was higher than, or similar to, the monotrait-heteromethod correlations. Finally, the heterotrait-heteromethod correlations were generally low and smaller than monotrait-monomethod and monotrait-heteromethod correlations, with the exception of Lab-TAB positive affect ($-.27$), which was greater in absolute value than the two observer-rated monotrait-heteromethod correlations.

Discussion

The majority of research on childhood BI relies on laboratory observational protocols. Whereas some researchers contend that observational methods are the optimal way of assessing temperament (Kagan, 1998), many maintain that the use of questionnaire measures can add unique and valuable information (e.g., Mangelsdorf et al., 2000; Rothbart & Bates, 2006). However, there are few rating scales specifically designed to assess BI, and research evaluating the psychometric properties of existing measures is limited. The Behavioral Inhibition Questionnaire (BIQ) offers an efficient method of assessing BI across different contexts and different informants. The present study sought to replicate the original report of the BIQ's psychometric properties (Bishop et al., 2003) and extend their findings by using a U.S. sample and a broader range of measures of convergent validity, and assessing discriminant validity.

Our findings are generally consistent with the initial report demonstrating satisfactory reliability and concurrent validity of the BIQ. We found that both the parent and teacher versions of the BIQ had acceptable internal consistency for total and subscale scores. Also, parent- and teacher-reported total BIQ scores were moderately correlated, although teachers rated children to be more inhibited in unfamiliar peer situation and performance situation than parents. Consistent with Bishop et al. (2003), there were no gender differences on either parent or teacher total BIQ scores. Also consistent with Bishop's data, we found that girls are more inhibited in physically challenging situations. However, our findings differed on several other subscales. Contrary to the prior study, both parents and teachers reported boys to be more inhibited in performance situations. In addition, we found that parents reported boys to be more inhibited in preschool/separation situations, and we did not replicate Bishop et al.'s finding that girls are more inhibited in unfamiliar adult situations.

Confirmatory factor analysis yielded a structure of six correlated factors loading onto a single general dimension of BI. The fit was acceptable for parent reports, but only marginal for teacher reports. BI responses clustered around the six contexts assessed by the BIQ subscales (approaching peers, approaching unfamiliar adults, performing in front of others, separation and preschool situations, unfamiliar situations in general, and novel physical activities suggestive of minor risks). This is consistent with previous studies suggesting that there may be some context specificity for BI (Kochanska, 1991; Rubin, Hastings, Stewart, Henderson, & Chen, 1997; Stevenson-Hinde, 1989). However, similar to Bishop et al.'s (2003) findings, the six first-order factors loaded significantly onto a single second-order factor, indicating that the BI can be considered a higher-order disposition that is expressed in multiple contexts. Taken together, these findings suggest that while driven by an overall tendency toward BI, the display of inhibited behaviors may vary somewhat as a function of context.

Although the six correlated factor model provided the best fit for the teacher ratings, the fit was at best marginal. Inspection of modification indices did not reveal ways to significantly improve model fit. The poorer fit for the teacher, compared to the parent, model may be due to the smaller N, resulting in less stable estimates. In addition, the preschool and daycare settings varied widely in terms of the length of time the teachers/caretakers knew the child, the number of hours of contact per week, and the number of other children in the classroom/daycare, hence the teacher ratings may have included greater measurement error.

Evidence for the convergent and discriminant validity of the BIQ was examined through the use of a multitrait-multimethod approach. Traits that are very closely related to BI (i.e. fear, shyness) as well as those related but conceptually distinct from BI (i.e. low PE, high NE and low constraint) were selected. These variables were assessed using a variety of methods, including parent and teacher questionnaires, a semi-structured diagnostic interview with a parent, behavioral coding of a standardized laboratory observation, and global ratings by an experimenter based on interactions with the child throughout the laboratory session.

Parent-reported BIQ showed relatively good convergent and discriminant validity. Fear and shyness as measured by parent questionnaires were used for monotrait-monomethod analyses. High correlations were found between the parent BIQ and the shyness scales from the CBQ (Rothbart, Ahadi, Hersey, & Fisher, 2001) and the CSPS (Coplan et al., 2004). Associations with the CBQ fear subscale was lower and in the moderate range, possibly because the behaviors reflected in the fear scale are not confined to novel situations/stimuli. With regard to monotrait-heteromethod correlations, the BIQ was moderately-highly correlated with interview-based ratings of BI. Given that the interview was based on parent report, the high correlation with the parent BIQ scores was not unexpected. Similar to Bishop et al. (2003), we found low-to moderate correspondence between BIQ scores and

behavioral ratings of the Lab-TAB episodes. In addition, we found a moderate association between parent BIQ ratings and experimenters' global ratings from the laboratory session. Our observational measures were more extensive than Bishop et al.'s (2003) in that their ratings were limited to a stranger-child interaction episode, whereas we also included several episodes in which the child was exposed to novel and mildly frightening or challenging stimuli and used observer's global ratings in addition to more molecular behavioral ratings. These correlations are consistent with the magnitudes of the associations between observational and parent-report measures of other temperament constructs (e.g., Majdandžić, van den Boom & Heesbeen, 2008).

Heterotrait-monomethod associations were generally smaller than the monotrait-monomethod and monotrait-heteromethod correlations. The largest heterotrait-monomethod correlation was with parent ratings of low PA on the CBQ smiling/laughter scale. As discussed earlier, BI and low PA share common features of low approach/exuberance, but differ in underlying motivations and eliciting contexts (Laptook, Klein, Olino, Dyson, & Carlson, 2010; Polak-Toste & Gunnar, 2006). Lastly, the heterotrait-heteromethod correlations were lower than almost all the other associations.

For teacher-reported BIQ, the monotrait-monomethod analysis revealed a high correlation with teacher reports of shyness on the CSPS. In the monotrait-heteromethod analyses, teacher BIQ showed low to moderate correlations with interview-based ratings, and somewhat lower associations with experimenters' global ratings and laboratory observation ratings of BI. Both heterotrait-monomethod and heterotrait-heteromethod correlations varied across traits. Most associations were smaller than the monotrait-monomethod and monotrait-heteromethod correlations, except for the moderate association with social disinterest. This may be due to the difficulty of distinguishing inhibited behavior from low desire for social interaction in a classroom/daycare setting. The core distinction between the two constructs lies in the underlying motivational state, rather than observable behaviors, making it difficult to discern in the absence of close and insightful monitoring. In sum, these results provide preliminary evidence for moderate to high convergent and discriminant validity for the parent and teacher versions of the BIQ.

Several limitations of this study should be considered. First, our sample consisted of predominantly middle-class, Caucasian, and non-clinical participants, limiting the generalizability of the results to different populations. Second, our findings are based on cross-sectional data. Future research with longitudinal designs can provide additional information to compare questionnaire and observational methods of assessing BI with regard to temporal stability and predicting the development of anxiety disorders. Third, as we did not obtain teacher reports on the full sample (either because the child was not in preschool/daycare or because the teacher did not complete the BIQ), and teachers' knowledge of the children varied, we may not have provided as strong a test of the validity of the teacher BIQ as of the parent BIQ. Lastly, while the use of experimenter ratings provided another source of observational data, our interpretations based on these ratings are limited by the low internal consistency and lack of interrater reliability data for this measure. Despite these shortcomings, the present study is noteworthy in that it adds to the limited data on the psychometric properties of the BIQ and extends the findings to a U.S. sample. Additionally, the use of diverse methods for assessing BI, including a semi-structured diagnostic interview as well as experimenter ratings following an extensive observational protocol, extends the evidence for the convergent validity of the BIQ and provides the first examination of the instrument's discriminant validity.

For over two decades, numerous studies have explored the nature of BI, particularly its neurobiological correlates, continuity and discontinuity over time, and implications for

psychopathology. However, a number of issues need to be addressed in order to expand our understanding of this construct, including further development and validation of assessment measures (Hirshfeld-Becker et al., 2008). Our findings extend Bishop et al.'s (2003) original study by providing further evidence for the reliability and validity of the BIQ as a measure of children's BI. The BIQ satisfies the need for a brief, economical measure that can be used to complement more expensive and time-consuming laboratory observations. This study adds to the small literature examining the convergence among different methods of assessing BI and provides further support for the view that parents, and to a more limited extent, teachers, can provide reliable and valid reports of preschool children's BI.

Acknowledgments

This work was supported by NIMH RO1 MH069942 (DNK) and GCRC Grant M01-RR10710 to Stony Brook University from the National Center for Research Resources.

References

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders 4th edition, text revision. Washington D.C.: American Psychiatric Association; 2000.
- Arbuckle, J.L. Amos (Version 7.0) [Computer Program]. Chicago: SPSS; 2006.
- Asendorpf JB. The malleability of behavior inhibition: A study of individual developmental functions. *Developmental Psychology*. 1994; 30:912–919.
- Bishop G, Spence SH, McDonald C. Can parents and teachers provide a reliable and valid report of behavioral inhibition? *Child Development*. 2003;1899–1917. [PubMed: 14669903]
- Broberg A, Lamb ME, Hwang P. Inhibition: Its stability and correlates in sixteen-to forty-month-old children. *Child Development*. 1990; 61:1153–1163. [PubMed: 2209185]
- Byrne, BM. Structural equation modeling with AMOS: Basic concepts, applications, and programming. Hillsdale, NJ: Erlbaum; 2001.
- Calkins SD, Fox NA, Marshall TR. Behavioral and physiological antecedents of inhibited and uninhibited behavior. *Child Development*. 1996; 67:523–540. [PubMed: 8625726]
- Campbell D, Fiske D. Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin*. 1959; 56:81–105. [PubMed: 13634291]
- Caspi, A.; Shiner, RL. Personality development. In: Damon, W.; Lerner, R.; Eisenberg, N., editors. *Handbook of child psychology: Vol. 3. Social, emotional, and personality development*. 6th ed. New York: Wiley; 2006. p. 300–364.
- Chen F, Bollen KA, Paxton P, Curran PJ, Kirby JB. Improper solutions in structural equation models. *Sociological Methods and Research*. 2001:468–508.
- Chronis-Toscano A, Degnan KA, Pine DS, Perez-Edgar K, Henderson HA, et al. Stable early maternal report of behavioral inhibition predicts lifetime social anxiety disorder in adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2009; 48:928–935. [PubMed: 19625982]
- Coplan RJ, Armer M. A multitude of solitudes: A closer look at social withdrawal and nonsocial play in early childhood. *Child Development Perspectives*. 2007; 1:26–32.
- Coplan RJ, Prakash K, O'Neil K, Armer M. Do you “want” to play? Distinguishing between conflicted shyness and social disinterest in early childhood. *Developmental Psychology*. 2004; 40:244–258. [PubMed: 14979764]
- Davidson RJ. Emotion and affective style: Hemispheric substrates. *Psychological Science*. 1992; 3:39–43.
- De Pauw SSW, Mervielde I. Temperament, personality and developmental psychopathology: A review based on the conceptual dimensions underlying childhood traits. *Child Psychiatry and Human Development*. 2010; 41:313–329. [PubMed: 20238477]
- Durbin EC, Hayden EP, Klein DN, Olino TM. Stability of laboratory assessed temperament traits from ages 3 to 7. *Emotion*. 2007; 7:388–399. [PubMed: 17516816]

- Durbin CE, Klein DN, Hayden EP, Buckley ME, Moerk KC. Temperamental emotionality in preschoolers and parental mood disorders. *Journal of Abnormal Psychology*. 2005; 114:28–37. [PubMed: 15709809]
- Egger, HL.; Ascher, BH.; Angold, A. The Preschool Age Psychiatric Assessment: Version 1.1. Durham, NC: Center for Developmental Epidemiology, Department of Psychiatry and Behavioral Sciences, Duke University Medical Center; 1999.
- Egger HL, Erkanli A, Keeler G, Potts E, Walter BK, Angold A. Test-retest reliability of the Preschool Age Psychiatric Assessment (PAPA). *Journal of the American Academy of Child & Adolescent Psychiatry*. 2006; 45:538–549. [PubMed: 16601400]
- Eisenberg N, Fabes RA, Guthrie IK, Murphy BC, Maszk P, Holmgren R, Suh K. The relations of regulation and emotionality to problem behavior in elementary school children. *Development and Psychopathology*. 1996; 8:141–162.
- Fox NA, Henderson HA, Rubin KH, Calkins SD, Schmidt LA. Continuity and discontinuity of behavioral inhibition and exuberance: Psychophysiological and behavioral influences across the first four years of life. *Child Development*. 2001; 72:1–21. [PubMed: 11280472]
- Goldsmith HH, Reilly J, Lemery KS, Longley S, Prescott A. Laboratory Temperament Assessment Battery: Preschool version. 1995 Unpublished manuscript.
- Gray, JA. The neuropsychology of temperament. In: Strelau, J.; Angleitner, A., editors. *Explorations in temperament. International perspectives on theory and measurement*. New York: Plenum Press; 1991. p. 105-128.
- Hirshfeld-Becker DR, Micco J, Henin A, Bloomfield A, Biederman J, Rosenbaum J. Behavioral inhibition. *Depression and Anxiety*. 2008; 25:357–367. [PubMed: 18412062]
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999; 6:1–55.
- Izard, CE.; Dougherty, LM.; Hembree, EA. A system for identifying facial expressions by holistic judgment (AFFEX). Newark: University of Delaware Media Services; 1989.
- Kagan, J. Biology and the child. In: Damon, W.; Eisenberg, N., editors. *Handbook of child psychology: Vol. 3. Social, emotional and personality Development*. 5th ed. New York: Wiley; 1998. p. 177-235.
- Kagan J, Reznick JS, Snidman N. Biological bases of childhood shyness. *Science*. 1988a; 240:167–171. [PubMed: 3353713]
- Kagan J, Reznick JS, Snidman N, Gibbons J, Johnson MO. Childhood derivatives of inhibition and lack of inhibition to the unfamiliar. *Child Development*. 1988b; 59:1580–1589. [PubMed: 3208569]
- Kagan, J.; Snidman, N. *The long shadow of temperament*. Cambridge, MA: Belknap Press; 2004.
- Kagan J, Snidman N, Kahn V, Towsley S. The preservation of two infant temperaments into adolescence. *Monographs of the Society for Research on Child Development*. 2007; 72:1–95.
- Kochanska G. Patterns of inhibition to the unfamiliar in children of normal and affectively ill mothers. *Child Development*. 1991; 62:250–263. [PubMed: 2055122]
- Kochanska G, Knaack A. Effortful control as a personality characteristic of young children: Antecedents, correlates, and consequences. *Journal of Personality*. 2003; 71:1087–1112. [PubMed: 14633059]
- Laptook RS, Klein DN, Olino TM, Dyson MW, Carlson G. Low positive affectivity and behavioral inhibition in preschool-age children: A replication and extension of previous findings. *Personality and Individual Differences*. 2010; 48:547–551. [PubMed: 20224804]
- Loehlin, JC. *Latent variable models: An introduction to factor, path and structural equation analysis*. 4th ed. Hillsdale, NJ: Erlbaum; 2004.
- Majdandžić M, van den Boom DC, Heesbeen DGM. Peas in a pod: Biases in the measurement of sibling temperament? *Developmental Psychology*. 2008; 1354–1368. [PubMed: 18793068]
- Mangelsdorf, SC.; Schoppe, SJ.; Buur, H. The meaning of parental reports: A contextual approach to the study of temperament and behavior problems in childhood. In: Molfese, VJ.; Molfese, DL., editors. *Temperament and personality development across the life span*. Mahwah, NJ: Lawrence Erlbaum Associates; 2000. p. 121-140.

- Marsh HM, Hocevar D. Application of confirmatory factor analysis to the study of self-concept: First- and higher-order factor models and their invariance across groups. *Psychological Bulletin*. 1985; 97:562–582.
- McManis MH, Kagan J, Snidman NC, Woodward SA. EEG asymmetry, power, and temperament in children. *Developmental Psychobiology*. 2002; 41:169–177. [PubMed: 12209658]
- Muris P, Meesters C, Spinder M. Relationships between child- and parent-reported behavioral inhibition and symptoms of anxiety and depression in normal adolescents. *Personality and Individual Differences*. 2003; 34:759–771.
- Perez-Edgar K, Roberson-Nay R, Hardin MG, Poeth K, Guyer AE, Nelson EE, et al. Attention alters neural responses to evocative faces in behaviorally inhibited adolescents. *Neuroimage*. 2007; 35:1538–1546. [PubMed: 17376704]
- Pfeifer M, Goldsmith HH, Davidson RJ, Rickman M. Continuity and change in inhibited and uninhibited children. *Child Development*. 2002; 73:1474–1485. [PubMed: 12361313]
- Polak-Toste, CP.; Gunnar, MR. Temperamental exuberance: Correlates and consequences. In: Marshall, PJ.; Fox, NA., editors. *The development of social engagement: Neurobiological perspectives*. Oxford: Oxford University Press; 2006. p. 19-45.
- Presley R, Martin R. Toward a structure of preschool temperament: Factor structure of the Temperament Assessment Battery for Children. *Journal of Personality*. 1994; 62:415–448. [PubMed: 7965566]
- Rapee RM, Kennedy S, Ingram M, Edwards S, Sweeney L. Prevention and early intervention with anxiety disorders in inhibited preschool children. *Journal of Consulting and Clinical Psychology*. 2005; 73:488–497. [PubMed: 15982146]
- Rothbart MK, Ahadi SA, Hersey KL, Fisher P. Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. *Child Development*. 2001; 72:1394–1408. [PubMed: 11699677]
- Rothbart, MK.; Bates, JE. Temperament. In: Damon, W.; Lerner, R.; Eisenberg, N., editors. *Handbook of child psychology: Vol. 3. Social, emotional, and personality development*. 6th ed. New York: Wiley; 2006. p. 99-166.
- Rubin KH, Hastings PD, Stewart SL, Henderson HA, Chen X. The consistency and concomitants of inhibition: Some of the children, all of the time. *Child Development*. 1997; 68:467–483. [PubMed: 9249961]
- Scarpa A, Raine A, Venables PH. The stability of inhibited/uninhibited temperament from ages 3 to 11 years in Mauritian children. *Journal of Abnormal Child Psychology*. 1995; 23:607–618. [PubMed: 8568083]
- Schmidt LA, Fox NA. Fear-potentiated startle responses in temperamentally different human infants. *Developmental Psychobiology*. 1998; 32:113–120. [PubMed: 9526686]
- Schwartz CE, Wright CI, Shin LM, Kagan J, Rauch SL. Inhibited and uninhibited infants “grown up”: Adult amygdalar response to novelty. *Science*. 2003; 300:1952–1953. [PubMed: 12817151]
- Stevenson-Hinde, J. Behavioral inhibition: Issues of context. In: Reznick, JS., editor. *Perspectives on behavioral inhibition*. Chicago: University of Chicago Press; 1989. p. 125-138.
- Zentner M, Bates JE. Child temperament: An integrative review of concepts, research programs, and measures. *European Journal of Developmental Science*. 2008; 2:7–37.

Figure 1.

The standardized parameter estimates for the second-order model (Model 4). Data for teacher report are presented in parentheses. Peer = Unfamiliar peers. Adults = Unfamiliar adults. Perf = Performance situations. Sepe = Separation/Preschool. Unfa = Unfamiliar situation. Phys = Physical challenges.

Table 1
Means, Standard Deviations and Coefficient Alpha for Mother and Teacher Ratings of Behavioral Inhibition

Factors	Parent ^a				Teacher ^b			
	M	SD	Gender Diff.	Alpha	M	SD	Gender Diff.	Alpha
Unfamiliar peer	Boys	18.15	8.33	F=3.20	19.78	7.79	F= 1.03	
	Girls	16.83	7.85	<i>p</i> = .07	18.64	9.13	<i>p</i> = .31	
	Total	17.54	8.13	<i>d</i> = .16	19.30	8.37	<i>d</i> = .14	.89
Unfamiliar adults	Boys	14.53	6.38	F= .95	15.45	6.45	F= .22	
	Girls	15.09	6.33	<i>p</i> = .33	15.03	7.02	<i>p</i> = .64	
	Total	14.79	6.35	<i>d</i> = .09	15.57	6.67	<i>d</i> = .06	.91
Performance situations	Boys	<u>12.60</u>	5.45	F= 9.63	<u>14.64</u>	6.10	F= 6.23	
	Girls	<u>11.15</u>	4.88	<i>p</i> < .01	<u>12.62</u>	5.84	<i>p</i> = .01	
	Total	11.93	5.24	<i>d</i> = .28	13.80	6.07	<i>d</i> = .34	.86
Separation/preschool	Boys	<u>12.58</u>	6.53	F= 9.64	11.55	5.70	F= .89	
	Girls	<u>10.84</u>	5.81	<i>p</i> < .01	10.82	5.80	<i>p</i> = .35	
	Total	11.78	6.26	<i>d</i> = .28	11.24	5.74	<i>d</i> = .13	.84
Unfamiliar situations	Boys	23.82	9.43	F= 1.95	17.49	7.32	F= 2.79	
	Girls	22.68	8.60	<i>p</i> = .16	15.81	7.67	<i>p</i> = .10	
	Total	23.30	9.06	<i>d</i> = .13	16.78	7.50	<i>d</i> = .22	.87
Physical challenge	Boys	<u>9.02</u>	4.19	F= 8.86	9.59	4.93	F= .13	
	Girls	<u>10.14</u>	4.20	<i>p</i> < .01	9.83	4.94	<i>p</i> = .72	
	Total	9.54	4.23	<i>d</i> = .27	9.69	4.92	<i>d</i> = .05	.82
Total behavioral inhibition	Boys	90.70	32.75	F= 2.15	88.69	31.39	F= 2.29	
	Girls	86.57	29.16	<i>p</i> = .14	82.07	33.27	<i>p</i> = .13	
	Total	88.79	31.18	<i>d</i> = .13	85.96	32.27	<i>d</i> = .21	.96

Note. Underlined entries signify significant gender differences (*p* < .05). *d* = effect size estimate of gender differences.

^a*N* = 495 (Total), *N* = 266 (Boys), and *N* = 229 (Girls).

^b $N = 226$ (Total), $N = 132$ (Boys), and $N = 94$ (Girls).

Table 2
Intercorrelations Between Subscales for Parent and Teacher Ratings of Behavioral Inhibition

	Parent-reported BIQ (N = 495)						
	1	2	3	4	5	6	7
1. Unfamiliar peer	.38	.29	.45	.32	.35	.04	.42
2. Unfamiliar adults	.44	.42	.46	.40	.46	.11	.52
3. Performance situations	.36	.30	.46	.40	.37	.05	.44
4. Separation/preschool	.26	.19	.28	.43	.27	.02	.33
5. Unfamiliar situations	.34	.31	.41	.42	.39	.08	.44
6. Physical challenge	.12	.20	.24	.24	.20	.09	.24
7. Total behavioral inhibition	.41	.34	.48	.45	.43	.07	.49

Note. Underline = $p < .01$. Bold = $p < .001$.

Table 3
Fit Indices for Parent and Teacher Reports of Behavioral Inhibition Questionnaire (BIQ)

Model	χ^2	Df	p	CFI	RMSEA	$\Delta\chi^2$	Δdf	$\Delta\chi^2/\Delta df$	p
Parent BIQ (N = 453)									
Model 1: 1 factor	3744.863	405	<.001	.674	.135 (.131 – .139)				
Model 2: 3 correlated factors	3065.008	402	<.001	.740	.121 (.117 – .125)	679.855	3	226.618	<.001
Model 3: 6 correlated factors	1382.606	390	<.001	.903	.075 (.071 – .079)	1682.402	12	140.200	<.001
Model 4: 6 first-order factors, 1 second-order factor	1424.966	400	<.001	.900	.075 (.071 – .080)			Target coefficient = .97	
Teacher BIQ (N = 199)									
Model 1: 1 factor	1453.396	350	<.001	.724	.126 (.119 – .133)				
Model 2: 3 correlated factors	1134.412	347	<.001	.803	.107 (.100 – .114)	318.984	3	106.328	<.001
Model 3: 6 correlated factors	822.673	335	<.001	.878	.086 (.078 – .093)	311.739	12	25.978	<.001
Model 4: 6 first-order factors, 1 second-order factor	865.619	344	<.001	.870	.088 (.080 – .095)			Target coefficient = .95	

Note. CFI = comparative fit index; RMSEA = root mean square error of approximation 90% confidence interval in parentheses

Table 4

Multitrait-multimethod Correlations for Parent-reported BIQ (N = 495)

	BIQ Total Score	PAPA BI	Lab-TAB BI	Exp. Rated BI	M	SD
BIQ Total Score	—				88.82	31.16
PAPA BI	.63***	—			3.39	3.72
Lab-TAB BI	.25***	.19***	—		.00	.40
Exp. Rated BI	.39***	.38***	.28***	—	2.90	.47
CBQ Fear	.30***	.18***	.26***	.16***	3.98	.94
CBQ Shyness	.85***	.64***	.23***	.44***	3.51	1.24
CSPS Shyness	.78***	.52***	.14**	.29***	1.99	.84
CBQ Anger/Frustration	.16***	.10*	.08	.04	4.57	.82
CBQ Smiling/Laughter	-.41***	-.24***	-.14**	-.12**	6.02	.53
CSPS Soc Disinterest	.31***	.15**	.04	.08	2.48	.77
CRS NE	.30***	.22***	.17***	.18***	19.57	4.78
Lab-TAB Sadness	.06	.09*	.25***	.07	1.34	.24
Lab-TAB Anger	-.05	-.04	.00	-.13**	1.35	.25
Lab-TAB PA	-.18***	-.18***	-.10*	-.22***	.41	.13
Lab-TAB impulsivity	-.18***	-.16***	-.13**	-.32***	.68	.33

Note. CBQ = Children's Behavior Questionnaire; CSPS = Child Social Preference Scale; PAPA = Preschool Age Psychiatric Assessment; CRS = Children's Reaction Scale; Lab-TAB = Laboratory Temperament Assessment Battery.

**
p < .01.

p < .001.

Table 5

Multitrait-multimethod Correlations for Teacher-reported BIQ (N = 226)

	BIQ Total Score	PAPA BI	Lab-TAB BI	Exp. Rated BI	M	SD
BIQ Total Score	–				85.67	32.29
PAPA BI	.46***	–			3.39	3.72
Lab-TAB BI	.15*	.19***	–		.00	.40
Exp. Rated BI	.25***	.38***	.28***	–	2.90	.47
CSPS Shyness	.72***	.32***	.05	.10	1.84	.83
CSPS Soc Disinterest	.45***	.30***	.05	–.01	2.69	1.07
CRS NE	.13	–.05	.00	–.19**	17.69	6.00
Lab-TAB Sadness	.04	.09*	.25***	.07	1.34	.24
Lab-TAB Anger	.06	–.04	–.01	–.13***	1.35	.25
Lab-TAB PA	–.27***	–.18***	–.10*	–.22***	.41	.13
Lab-TAB impulsivity	–.13*	–.16***	–.13***	–.32***	.68	.33

Note. CBQ = Children's Behavior Questionnaire; CSPPS = Child Social Preference Scale; PAPA = Preschool Age Psychiatric Assessment; CRS = Children's Reaction Scale; Lab-TAB = Laboratory Temperament Assessment Battery.

* $p < .05$.

** $p < .01$.

*** $p < .001$.