Child Development, November/December 2003, Volume 74, Number 6, Pages 1899–1917

Can Parents and Teachers Provide a Reliable and Valid Report of Behavioral Inhibition?

Gillian Bishop, Susan H. Spence, and Casey McDonald

Reliability and validity of parent and teacher report of behavioral inhibition (BI) was examined among children aged 3 to 5 years. Confirmatory factor analysis supported 6 correlated factors reflecting specific BI contexts, each loading on a single, higher order factor of BI. Internal consistency was acceptable, with moderate stability over 1 year and strong correlation with a brief inhibition subscale from a temperament questionnaire. Children who were rated by mothers and teachers as high BI took longer to initiate contact with a stranger, spoke less often and for shorter periods, and required more prompting to elicit speech compared with low-BI peers in a simulated stranger interaction task. Father report of BI was significantly associated with mean duration of speech and eye gaze.

Kagan, Reznick, Clarke, Snidman, and Garcia Coll (1984) coined the term *behavioral inhibition* (BI) to describe a relatively consistent pattern of behavioral and emotional responses to unfamiliar or novel people, places and situations, or objects. Inhibited children typically respond to novel situations with initial restraint, caution, low rates of approach, and quiet withdrawal, and with unfamiliar people they are usually shy, timid, and reticent (Garcia Coll, Kagan, & Reznick, 1984). Children who have high BI are more likely than their uninhibited peers to show a particular pattern of physiological features, including greater activation of the hypothalamic-pituitary axis, the reticular activating system, and the sympathetic arm of the autonomic nervous system in response to novelty (Kagan, Reznick, & Snidman, 1987). However, these physiological parameters are only evident for around one third of children identified as showing elevated BI and the exact brain mechanisms are yet to be identified (Kagan, Snidman, Arcus, & Reznick, 1994; Kagan, Snidman, & Arcus, 1998). Furthermore, Kagan et al. (1994) suggested that the physiological correlates and predictive stability of the trait are only evident for children at the extreme end of the BI profile. These authors suggest that for this reason BI reflects a categorical, qualitatively distinct construct involving a discrete pattern of behaviors and physiology resulting from a specific combination and configuration of genes. Other authors have challenged this view, suggesting that BI reflects a continuum of behavioral features, representing quantitative rather than qualitative differences in a manner similar to other temperament characteristics (Woodward, Lenzenweger, Kagan, Snidman, & Arcus, 2000). Although the debate and search for biological markers for the construct continue, Kagan et al. pointed out that identification of BI relies on the behavioral and affective profile of the child. Thus, the development of reliable, valid, and efficient methods of assessing the behavioral manifestations of BI is important.

One of the reasons BI warrants attention is the potential impact on children's psychosocial development. For example, BI children are at increased risk for developing internalizing disorders later in childhood (e.g., Biederman et al., 1993; Hirshfeld et al., 1992; Kagan, Snidman, Zentner, & Peterson, 1999) and adolescence (e.g., Prior, Smart, Sanson, & Oberklaid, 2000; Schwartz, Snidman, & Kagan, 1999). Clearly, there is a need for reliable and valid screening instruments for the assessment of BI on a large-scale basis to identify and intervene early with children who show this temperament pattern and who are at risk for its adverse consequences.

Different researchers have emphasized different methods of assessing BI. For example, in both their longitudinal and cross-sectional research, Kagan and his colleagues (see Kagan et al., 1994 for a Summary) identified BI through observation of children's behavior in laboratory-based tasks. In contrast, some researchers such as Asendorpf (1990a) have used naturalistic observations and parent reports of child behavior as indicators of BI. There are clear advantages and disadvantages of the

Gillian Bishop, Susan H. Spence, and Casey McDonald, School of Psychology, University of Queensland.

Correspondence concerning this article should be addressed to Susan H Spence, School of Psychology, University of Queensland, Brisbane, QLD 4072, Australia. Electronic mail may be sent to s.spence@psy.uq.edu.au.

different methods of assessment of this construct in terms of reliability, validity, cost, and feasibility.

Whether parent report can be used to assess BI or whether laboratory-based assessment is necessary has generated debate (e.g., Kagan, 1992). Those who argue against the use of parent report as an indicator of BI claim that parents may not be aware of much of their child's behavior unless it concerns highly salient events or interactions (Maccoby & Martin, 1983) and that parent report may be influenced by inaccurate recall or memory bias (Robins, 1963). However, many of the reported correlations between maternal report (maternal Q-Sort and interview) and behavioral and physiological measures of inhibition are relatively high. For example, Garcia Coll et al. (1984) found significant positive correlations between the laboratory-based behavioral observations of inhibition and mother ($r = .54$) and father ($r = .49$) reports of BI. Likewise, Kagan et al. (1984) found that maternal Q-Sorts assessing inhibited and uninhibited behavior were significantly related to the child's behavior in peer play situations in the expected direction (correlations ranged from .42 to .52). The maternal Q-Sorts also demonstrated good predictive validity, with the index of inhibition at 21 months being significantly correlated with the maternal Q-Sort at 4 years of age ($r = .59$ and .56 for inhibited and uninhibited children, respectively). Similarly, maternal interviews at 21 months predicted the index of inhibited behavior at ages 4 ($r = .53$) and 5.5 ($r = .53$; Reznick et al., 1986).

Proponents of using parent report to assess child BI point out that this method has the advantage of being simple and easy to administer and score, thereby making it a fast and economical method of obtaining information. In contrast, behavioral observation in laboratory settings tends to be a costly and time-consuming exercise, being cumbersome to construct and administer and typically yielding hours of videotaped footage that must be coded by experienced and trained staff (Tryon, 1998). A further advantage of parent report is that parents have generally observed their child over prolonged periods across a broad range of real-life situations (DiLalla & Falligant, 1995) and can thus provide a fuller picture than can be obtained from a relatively brief laboratory task involving restricted situations. In an attempt to increase the validity of observational data, some researchers have attempted to assess BI in naturalistic settings (Asendorpf, 1990a). This approach also has its disadvantages in that human observers or cameras may influence the child's behavior. There are also difficulties in tracking fast-moving children who tend to disappear out of sight or out of earshot. Moreover, observers can be biased by their expectations and prejudices, and this may influence their recording and decoding of behavior (Tryon, 1998).

Thus, neither behavioral observation nor parent report methods are without their limitations. However, given the importance of early detection there remains a need for inexpensive, rapid, and easy-to-administer methods of assessing BI in children. At the time this study was developed, there was a paucity of reliable and valid parent or teacher questionnaires for the assessment of BI. Existing parent report measures were not considered satisfactory in that they were typically subscales of more global assessments of childhood temperament, reflecting a restricted range of contexts, such as the California Child Q-Sort (Block, 1978; Block & Block, 1980), the Preschool Characteristics Questionnaire (Finegan, Niccols, Zacher, & Hood, 1989), the Middle Childhood Temperament Questionnaire (Hegvik, McDevitt, & Carey, 1982), the Colorado Childhood Temperament Inventory (Rowe & Plomin, 1977), the Child Temperament Questionnaire (Thomas & Chess, 1977), and the Teacher Temperament Questionnaire (Thomas & Chess, 1977).

A further consideration in the development of the scale was the need to examine the child's behavior across a range of novel situations. There is evidence that BI is not a unitary construct and the degree of BI may vary across different contexts. For example, Kochanska (Kochanska, 1991; Kochanska & Radke-Yarrow, 1992) and Rubin, Hastings, Stewart, Henderson, and Chen (1997) identified different contexts in which inhibited behavior may occur, namely, unfamiliar settings, unfamiliar adults, and unfamiliar peers. Kagan and colleagues (e.g., Reznick et al., 1986) also found evidence of inhibited behavior in response to novel objects and physical activities that possessed an element of risk. Thus, in the assessment of BI it is important to examine child behavior across a range of contexts in keeping with the evidence outlined earlier.

Accordingly, the first stage of this study involved the development of a parent and teacher questionnaire that assessed children's BI across different contexts. Items were selected to reflect child behavior across six contexts reflecting three domains: response to social novelty (unfamiliar adults, peers, and performing in front of others), situational novelty (preschool and unfamiliar situations), and novel physical activities suggestive of minor risk. The study then examined whether parents and teachers could provide a reliable and valid assessment of children's BI. Direct behavioral observation

of children in a simulated, novel, stranger interaction situation was used to examine the construct validity of parent and teacher report. Parent and teacher report of BI was compared with direct observation of specific social behaviors that have been shown to be indicative of BI by other researchers in the field, including latency of initiation of conversation, frequency of speech, length of utterance, and total eye gaze (Asendorpf, 1990a, 1990b, 1992, 1993a, 1993b, 1994; Gersten, 1989; Hill, Lowers, Locke, Snidman, & Kagan, 1999). These behaviors are also consistent with the criteria used by Kagan and colleagues (Garcia Coll et al., 1984; Kagan et al., 1984; Kagan, Reznick, Snidman, Gibbons, & Johnson, 1988; Reznick et al., 1986) to categorize children into inhibited or uninhibited groups.

Method

Generation of Questionnaire Items and Pilot Work for the Behavioral Inhibition Questionnaire

An initial list of 40 items was developed from a review of existing literature relating to the features of BI across a range of contexts relevant to children (e.g., Asendorpf, 1990b, 1992, 1993a, 1993b,1994; Garcia Coll et al., 1984; Kagan et al., 1984, Kagan et al., 1988; Reznick et al., 1986; Snidman, 1989) and existing temperament questionnaires, including the Parental Inhibition Questionnaire (Asendorpf, 1987), California Child Q-Sort (Block, 1978; Block & Block, 1980), the Preschool Characteristics Questionnaire (Finegan et al., 1989), the Middle Childhood Temperament Questionnaire (Hegvik et al., 1982), the Colorado Childhood Temperament Inventory (Rowe & Plomin, 1977), the Child Temperament Questionnaire (Thomas & Chess, 1977), and the Teacher Temperament Questionnaire (Thomas & Chess, 1977). Several items were discarded following piloting if they were not readily understood by parents or were psychometrically unsound. This left 30 items in the parent version and 28 items in the teacher version. The items spanned six contexts reflecting three domains: social novelty (unfamiliar adults, 4 items; peers, 6 items; performance situations, 4 items), situational novelty (separation and preschool, 4 items; unfamiliar situations in general, 8 items), and physical activities in which there is a minor but possible risk of injury (4 items). Two items were not included in the teacher version as these related to the child's behavior in unfamiliar homes where teachers would be unlikely to observe the child. Each item was randomly allocated within the questionnaire. To prevent response bias, 16 of the items were worded in a positive way (e.g., ''happily approaches new situations or activities'') and 14 were worded in a negative way (e.g., ''approaches new situations or activities very hesitantly'').

The list of 30 items was then given to two experienced clinical child psychologists who specialize in BI. They were asked to place the items into the three domains of social novelty, situational novelty, and situations involving minor physical risk to which they thought each belonged. Their allocation of items into each of the three factors was virtually identical to that hypothesized by the researcher, except one item—''is independent.'' There was some disagreement as to whether this item tapped social, situational, or physical novelty. This item was allocated to the situational novelty domain (general context). The draft questionnaire was then administered to 10 sets of parents (i.e., mothers and fathers) and 5 teachers, who were asked for their feedback regarding the readability and understandability of instructions and questionnaire items. Minor changes were made to the wording of the items based on this pilot work before it was administered to the full sample. Parents and teachers were asked to rate, from 1 (*hardly ever*) to 7 (*almost always*), how frequently the behavior described in each item occurred for the target child. Copies of the scale may be obtained from the second author on request.

Participants

The parent sample was composed of 613 mothers (M age = 34.55 years, SD = 4.59) and 506 fathers (M age = 37.28 years, SD = 5.50) who completed the questionnaire about their child. The children were between 3 and 5 years of age (M = 48.99 months, SD = 5.92), with 50.4% girls (n = 312) and 49.6% boys (n = 307). In addition, 585 teachers of the children who were participating in the study completed the questionnaires.

The children were recruited from kindergartens and preschools in the Brisbane metropolitan area and were selected to encompass the spectrum of socioeconomic status and ethnic mix that is representative of the general Australian population. The majority of the parents in the sample were married (87.15%) and born in Australia (78%). Parents were employed across the full range of Australian Standard Classification of Occupations (Australian Bureau of Statistics, 1997) categories, but families in the upper middle to higher socioeconomic status brackets were overrepresented (41.55%) compared with the Australian population in general. Similarly,

although children came from a wide variety of ethnic backgrounds, most were of Caucasian origin (92%). Aside from socioeconomic status, the sample characteristics were representative of the general Australian population.

Families were excluded if the parents were unable to read English and if mental retardation or language disorder had been formally assessed and identified in the child. Written informed consent was obtained from parents before participation in the study.

To assess the stability of the Behavioral Inhibition Questionnaire (BIQ), the children were reassessed 1 year later. A total of 481 mothers (79%) and 381 fathers (75%) were contactable and returned completed questionnaires. The demographic characteristics of the sample 1 year later did not differ significantly from the original sample. The stability of teacher report was examined for 431 children, but the children typically had different teachers by the 12-month follow-up assessment.

To examine the concurrent validity of the parent and teacher questionnaires, the data were compared with indicators of BI obtained from direct behavioral observation. One hundred child participants (approximately 16% of the total subject pool) were randomly selected (from the full sample of 619 participants) to participate in the behavioral observations. The personal characteristics of children and parents in this subsample closely matched those outlined earlier for the full sample.

Measures

To assess its convergent validity, the parent and teacher forms of the BIQ were compared with the inhibition subscale of the Temperament Assessment Battery for Children–Revised (TABC–R; Presley & Martin, 1994), which has both parent and teacher forms. The TABC–R inhibition subscale consists of 10 items (e.g., ''My child is shy with adults he/she does not know'') that assess inhibited behavior. The scale is reported to have high internal consistency for both parent and teacher forms (R. P. Martin, personal communication, November 27, 1995; Ball, Pelco, Havill, & Reed, 2001). However, the TABC–R does not adequately assess the three contexts in which inhibition may be displayed and it predominantly focuses on inhibition in response to unfamiliar adults and peers (i.e., 7 of the 10 items focus on social inhibition), with only 3 items assessing inhibition in unfamiliar situations and no items assessing inhibition in response to novel physical activities that are suggestive of minor risk. Nevertheless, it was selected to examine the concurrent validity of the new measure as it is one of the few parent report questionnaires to assess the temperamental dimension of BI available for which data regarding its psychometric properties have been examined. In the present study the internal consistency was found to be acceptable for mother (alpha = .87) and father (alpha = .85) but unacceptable for teacher report (alpha = .60).

Procedure for Behavioral Observation

Behavioral observations, in which the children interacted with an adult stranger, were conducted and videotaped within the kindergarten or preschool using an adaptation of Asendorpf's (1987) version of the stranger situation. This component of the study was conducted for a random sample of 100 children as outlined previously.

Preobservation: Introduction of child to the observation. Outside the observation area, before the observation began, the teacher was instructed to say to the child the following: ''There are some really good toys in here. You can go in and play with them if you like.'' If the child refused to go into the area voluntarily, the teacher escorted the child into the area and showed the child the toys on the floor (i.e., basic building blocks). The teacher then said to the child, ''I'll be back in a little while. You stay here and play with these toys.'' The teacher then left the observation area. Approximately 2 min later, the stranger entered the area. The stranger was one of two young adult females dressed in smart-casual clothing.

Preinteraction phase. The stranger walked directly into the room and sat on a chair approximately 2 m from the child. As she sat down, the stranger briefly greeted the child and placed a bag of toys on the floor. After 5 s, the stranger inspected the toys without unpacking the bag, then briefly looked at the child with a slight smile three times. After 30 s the stranger slowly unpacked the bag, placing the toys on the floor in front of her, then looked around the area for 1 min. During the next 2 min she closely inspected and played with each toy in turn. The toys were a plastic dinosaur, a cash register, and a picture-viewing box (i.e., a series of interesting cartoon pictures could be scrolled onto the screen and specific parts of each picture could be viewed with an attached magnifying lens). The toys were inspected in this order (i.e., they aroused increasing interest in the child). Only contact initiation latency (i.e., the time the child took to initiate verbal contact with the stranger) was assessed in the preinteraction phase.

Interaction phase. In the event that the child did not initiate verbal contact with the stranger during the 3-min preinteraction phase, the stranger prompted the child using standard prompt questions. As soon as the child started talking to the stranger, or when the 3-min preinteraction phase was over, the stranger commenced a 3-min standardized interaction phase with the child. The response of the stranger varied in a preset format, depending on whether the child had initiated the verbal contact. If the child had initiated verbal contact, the stranger responded in an appropriate manner (i.e., the stranger answered any questions concisely in a friendly way) but did not ask the child any reciprocal questions or make any unsolicited comments to the child. If the child then made no further attempts at verbal contact, after a 15-s delay (starting from the end of the stranger's response to the child's utterance) the stranger prompted the child with a question from a sequence of 18 prompt questions. If the child responded verbally, the stranger let the child speak freely. If the stranger prompted the child about a particular topic area, but the child chose to discuss a different topic altogether, the stranger responded in an appropriate manner. Following the child's response, the stranger then waited 15 s before asking the next prompt question. This process was repeated using the list of prompt questions until the 3-min interaction phase was completed. After the child had been engaged in the conversation period for the required 3 min, the stranger ended any conversation with the child in an appropriate manner, thanked the child, and escorted him or her back to the class.

Coding and Analysis of the Stranger–Child Interaction

Both the phase before contact initiation (0–3 min) and the first 3 min of the stranger–child interaction were videotaped with one camera, which had a view of the child and the stranger, with the focus being on the child's face. This interaction sequence was subsequently coded for the following variables: contact initiation latency, frequency and duration of spontaneous comments or utterances to stranger during interaction phase, frequency and duration of child's silence during interaction phase, and frequency and duration of child's eye gaze toward the stranger during interaction phase. These observational variables were selected on the basis of research that indicated that they were sensitive in discriminating between behaviorally inhibited and uninhibited children (e.g., Asendorpf, 1987; Kagan et al., 1988; Reznick et al., 1986) and followed definitions used by Asendorpf (before 1987).

The time the child took to initiate verbal contact with the stranger served as the contact initiation latency (CIL). If the child did not initiate verbal contact within 3 min, CIL was defined as 180 s. Speech between the child and the stranger and eye gaze of the child toward the stranger during the first 3 min of the interaction phase were coded as follows.

The trained observer decoded the videotapes using a box fitted with two buttons (one coding speech, the other coding eye gaze) attached to a computer to record the amount of time (in milliseconds) the child spent speaking to and looking at the stranger. The computer calculated frequency of speech, total duration of speech, mean duration of speech utterances, and frequency and duration of silence periods.

Likewise, the rater pressed the eye gaze button for the duration of each of the child's periods of eye gaze toward the stranger. The computer program then automatically calculated frequency and total duration of eye gaze, mean duration of eye gaze episodes, and frequency and duration of looking away from the stranger.

Observer Rating of BI

The observer then rated the child's behavior during the stranger–child interaction on a 7-point scale for intensity of BI ranging from 1 (*extremely uninhibited*; e.g., initiates verbal contact quickly, appears confident when talking to stranger, talks a lot, requires no prompting) to 7 (*extremely inhibited*; e.g., does not talk at all, frequent and long gazes at stranger, fearful facial expression, might cry). The coders were not privy to the questionnaire data regarding the BI status of the children before rating them.

Training of Raters for Interrater Reliability

The two raters completed a total of 10 hr training. Both raters were provided with written descriptions and videotaped examples of the behaviors to be coded. The raters then coded a number of sample interactions and compared their coding with a criterion code for the same sample. Differences between the coding of the raters and the criterion code were discussed until both raters were clear about how and why their coding differed from the criterion.

Interrater Reliability for Stranger–Child Interaction Coding

One researcher coded all of the interactions. Interrater reliability was checked using the second

trained observer, who independently coded a random sample of 15% of the interactions. The manner in which the data were stored by the computer during the decoding process provided only total scores (i.e., frequency and duration of speech, frequency and duration of eye gaze) for the 3-min observation period and did not provide moment-by-moment information about specific behaviors. Thus, interrater reliability was calculated using Pearson's product–moment correlation and single-measure, two-way (judges fixed) effect model intraclass correlation (ICC) coefficients. Interrater reliability exceeded $r = .90$, ICC $= .90$, 95% confidence interval (CI) = .73 to .97, for all behavioral categories other than frequency of speech, $r = .71$, ICC = .67, 95% CI = .27 to .88.

Results

Confirmatory Factor Analysis

Confirmatory factor analysis was conducted for mother, father, and teacher report separately to determine whether factor structure was consistent with the constructs proposed to underpin the measure. The data were examined using EQS (Bentler, 1985) with maximum likelihood (ML) estimation using the correlation matrix.

The purpose of structural equation modeling is to determine the extent to which the hypothesized model fits, or adequately describes the sample data (Byrne, 1994). In EQS, model fit is assessed by examining the degree of similarity between the sample covariance matrix and the restricted (or predicted) covariance matrix. Discrepancy in fit between these matrices is represented by the residual covariance matrix and its standardized version, which takes standard deviations on the measured variables into account (Byrne, 1994). The EQS program (Bentler, 1985) produces a range of goodness of fit indexes. The independence chi-square is a likelihood ratio test statistic that evaluates the fit between the restricted hypothesized model and the unrestricted sample data. The model may be rejected if the chi-square value is large relative to the degrees of freedom and accepted if the value is nonsignificant or small. However, for very large sample sizes there is a risk of relatively good-fitting models being rejected on the basis of the chi-square test (Marsh, 1994; Marsh, Balla, & McDonald, 1988). Therefore, the fit of the model should be interpreted on a range of statistics, such as the non-normed fit index (NNFI), comparative fit index (CFI), root mean square residual (RMR), and root mean squared error of approximation (RMSEA). The CFI and the NFI provide a relatively nonbiased indication of fit for large sample sizes (Gerbing & Anderson, 1993; McDonald & Marsh, 1990). Values for CFI and NNFI greater than .90 are generally regarded as representing an acceptable fit of the model to the data (Gerbing & Anderson, 1993). The RMR is an index of the degree of discrepancy between elements in the sample and the hypothesized covariance matrix. If there is a good fit between the hypothesized model and the sample, the RMR will be small, with a good fit reflecting an RMR close to .05 or lower (possible values range from 0 to 1.00). The RMSEA provides a measure of degree of discrepancy per degrees of freedom. Browne and Cudeck (1993) suggested that an RMSEA value of .05 or lower reflects a close fit between the hypothesized model and the sample data. In the present data set there was evidence of significant kurtosis for several items for all informants ($p < .001$). Attempts to transform items or to use other forms of estimation using elliptical or arbitrary distribution theories did not change the pattern of results; thus, the analyses reported here are based on nontransformed data using ML extraction. However, to take into account deviation from normality, the Satorra and Bentler (1988) test statistic was included in the examination of model fit.

For mothers, fathers, and teachers, four models were compared: (a) single factor, (b) three correlated factors, (c) six correlated factors, and (d) six correlated factors loading onto one higher order factor. The first comparison model (Model 1) was a single-factor model in which all questionnaire items can be regarded as assessing a single, homogenous dimension of inhibition. In such a model, the data are best explained by a single factor onto which all dimensions of inhibition load strongly, with minimal variance left to be explained by the different contexts of inhibition. However, given the proposition that BI is not a unitary construct and the degree of BI may vary across different contexts (Kagan et al., 1998), we examined the possibility that the data would be better explained by a model in which the items of the questionnaire cluster around either the three domains in which BI can occur (Model 2: social novelty, situational novelty, and novel physical activities involving minor risk) or the six contexts that lay within these three domains (Model 3: unfamiliar adults, peers, performing in front of others, separation and preschool, unfamiliar situations in general, and novel physical activities involving of minor risk). It was assumed that in both Models 2 and 3 the factors would be intercorrelated. Finally, confirmatory factor analysis was used to determine whether

Table 1
Fit Indexes for Each Model for Mother, Father, and Teacher Behavioral Inhibition Questionnaire (BIQ)

Model	$\chi 2$	*df*	*p*	S-B $\chi 2$	NNFI	CFI	S-B	RMR	RMSEA	Comparison	$\chi 2$ change	*df* change	*P* of $\chi 2$ change
						Mother BIQ							
Null model	13287	435											
Model 1: 1 factor	4494	405	<.001	3635	.66	.68	.69	.09	.13	Null vs. 1	8793	30	<.001
Model 2: 3 correlated factors	3358	402	<.001	2718	.75	.77	.78	.08	.11	2 vs. 1	1136	3	<.001
Model 3: 6 correlated factors	1835	390	<.001	1480	.87	.88	.89	.06	.08	3 vs. 2	1532	12	<.001
Model 4: 6 first-order factors, 1 second-order factor	1910	400	<.001	1541	.87	.88	.89	.06	.08	Target coefficient = .96			
						Father BIQ							
Null model	9169	435		435									
Model 1: 1 factor	3232	405	<.001	2517	.65	.68	.71	.08	.12	Null vs. 1	5937	30	<.001
Model 2: 3 correlated factors	2632	402	<.001	2055	.72	.75	.77	.08	.11	2 vs. 1	600	3	<.001
Model 3: 6 correlated factors	1569	390	<.001	1245	.85	.87	.88	.06	.08	3 vs. 2	1063	12	<.001
Model 4: 6 first-order factors, 1 second-order factor	1633	400	<.001	1293	.85	.86	.88	.06	.08	Target coefficient = .96			
						Teacher BIQ							
Null model	15354	378											
Model 1: 1 factor	3446	350	<.001	2473	.78	.79	.85	.07	.12	Null vs. 1	11908	28	<.001
Model 2: 3 correlated factors	2236	347	<.001	1669	.86	.87	.90	.06	.10	2 vs. 1	1210	3	<.001
Model 3: 6 correlated factors	1520	335	<.001	1109	.91	.92	.94	.05	.08	3 vs. 2	716	12	<.001
Model 4: 6 first-order factors, 1 second-order factor	1743	344	<.001	1263	.89	.90	.93	.05	.08	Target coefficient = .87			

Note. S-B = Satorra-Bentler fit index; NNFI = non-normed fit index; CFI = comparative fit index; RMR = root mean square residual; RMSEA = root mean square error of approximation.

the covariance between the first-order contextual factors of BI could be satisfactorily explained by a single higher order factor reflecting BI in general (Model 4).

Model 1 (single factor). For mother report, all items loaded significantly on the single factor, with 28 of the 30 BIQ items having loadings greater than .40. For father report, again all items loaded significantly on the single factor, with 27 of the 30 BIQ items having loadings greater than .40. On the teacher BIQ, all 28 BIQ question items loaded significantly on the single factor, with loadings greater than .50. However, Table 1 indicates that the one-factor solution did not provide a good fit of the data, with fit indexes below .80 for all three informants.

Model 2 (three correlated factors). All items loaded greater than .50 on their hypothesized factors for all three informants. Comparison with Model 1 indicated that the chi-square changed in relation to change in the degrees of freedom, suggesting that the three-correlated-factor model (Model 2) provided a significantly better fit than the one-factor model. However, the fit indexes were still relatively poor, suggesting that overall Model 2 did not provide a satisfactory fit of the data.

Model 3 (six correlated factors). All items loaded greater than .50 on their hypothesized factors for all three informants as shown in Table 2. However, as indicated in Table 1, Model 3 provided a significantly better fit of the data than did Model 2 in terms of change in chi-square statistic in relation to change in the degrees of freedom. The NNFI, CFI, and Satorra-Bentler fit indexes of Model 3 were relatively good, with values approaching .90 and RMR values

Table 2
Confirmatory Factor Analysis Loadings of Inhibition Items on Three Contextual Factors: Mother, Father, and Teacher Behavioral Inhibition Questionnaire (BIQ)

Predicted factor	Questionnaire item	F1 Mo	F2 Mo	F3 Mo	F4 Mo	F5 Mo	F6 Mo	F1 Fa	F2 Fa	F3 Fa	F4 Fa	F5 Fa	F6 Fa	F1 Te	F2 Te	F3 Te	F4 Te	F5 Te	F6 Te
Peers	Will happily approach a group of unfamiliar children and join in their play	.81	–	–	–	–	–	.81	–	–	–	–	–	.80	–	–	–	–	–
	Is comfortable asking other children to play	.75	–	–	–	–	–	.72	–	–	–	–	–	.80	–	–	–	–	–
	Is shy when first meeting new children	.82	–	–	–	–	–	.76	–	–	–	–	–	.85	–	–	–	–	–
	Is reluctant to approach a group of unfamiliar children to ask to join in	.76	–	–	–	–	–	.76	–	–	–	–	–	.75	–	–	–	–	–
	Is very friendly with children he or she has just met	.77	–	–	–	–	–	.73	–	–	–	–	–	.80	–	–	–	–	–
	Tends to watch other children, rather than join in their games	.72	–	–	–	–	–	.71	–	–	–	–	–	.77	–	–	–	–	–
Physical challenges	Is cautious in activities that involve physical challenge (e.g., climbing, jumping from heights)	–	.75	–	–	–	–	–	.53	–	–	–	–	–	.74	–	–	–	–
	Is confident in activities that involve physical challenge (e.g., climbing, jumping from heights)	–	.87	–	–	–	–	–	.70	–	–	–	–	–	.82	–	–	–	–
	Is hesitant to explore new play equipment	–	.58	–	–	–	–	–	.55	–	–	–	–	–	.80	–	–	–	–
	Happily explores new play equipment	–	.66	–	–	–	–	–	.78	–	–	–	–	–	.78	–	–	–	–
Separation/ Preschool	Happily separates from parent(s) when left in new situations for the first time (e.g., kindergarten, preschool, child care)	–	–	.85	–	–	–	–	–	.76	–	–	–	–	–	.76	–	–	–
	Quickly adjusts to new situations (e.g., kindergarten, preschool, child care)	–	–	.82	–	–	–	–	–	.81	–	–	–	–	–	.89	–	–	–
	Gets upset at being left in new situations for the first time (e.g., kindergarten, preschool, child care)	–	–	.83	–	–	–	–	–	.76	–	–	–	–	–	.74	–	–	–

	Takes many days to adjust to new situations (e.g., kindergarten, preschool, child care)	–	–	.87	–	–	–	–	–	.82	–	–	–	–	–	.85	–	–	–
Performance situations	Enjoys being the center of attention	–	–	–	.55	–	–	–	–	–	.52	–	–	–	–	–	.81	–	–
	Is happy to perform in front of others (e.g., singing, dancing)	–	–	–	.83	–	–	–	–	–	.81	–	–	–	–	–	.87	–	–
	Dislikes being the center of attention	–	–	–	.67	–	–	–	–	–	.61	–	–	–	–	–	.90	–	–
	Is reluctant to perform in front of others (e.g., singing, dancing)	–	–	–	.84	–	–	–	–	–	.85	–	–	–	–	–	.82	–	–
Adults	Is very quiet around new (adult) guests to our home (preschool for teacher version)	–	–	–	–	.78	–	–	–	–	–	.73	–	–	–	–	–	.88	–
	Is very talkative to adult strangers	–	–	–	–	.86	–	–	–	–	–	.84	–	–	–	–	–	.91	–
	Happily chats to new (adult) visitors to our home (preschool for teacher version)	–	–	–	–	.86	–	–	–	–	–	.86	–	–	–	–	–	.92	–
	Is very quiet with adult strangers	–	–	–	–	.88	–	–	–	–	–	.82	–	–	–	–	–	.91	–
Unfamiliar situations	Approaches new situations or activities very hesitantly	–	–	–	–	–	.74	–	–	–	–	–	.65	–	–	–	–	–	.82
	Settles in quickly when we visit the homes of people we don't know well	–	–	–	–	–	.69	–	–	–	–	–	.70	–	–	–	–	–	–
	Is independent	–	–	–	–	–	.55	–	–	–	–	–	.48	–	–	–	–	–	.59
	Seems comfortable in new situations	–	–	–	–	–	.85	–	–	–	–	–	.80	–	–	–	–	–	.87
	Is clingy when we visit the homes of people we don't know well	–	–	–	–	–	.74	–	–	–	–	–	.70	–	–	–	–	–	–
	Happily approaches new situations or activities	–	–	–	–	–	.81	–	–	–	–	–	.76	–	–	–	–	–	.84
	Is outgoing	–	–	–	–	–	.78	–	–	–	–	–	.67	–	–	–	–	–	.80
	Seems nervous or uncomfortable in new situations	–	–	–	–	–	.75	–	–	–	–	–	.73	–	–	–	–	–	.81

Note. Mo = mother report; Fa = father report; Te = teacher report.

approaching .05. Thus, for mothers, fathers, and teachers, there is support for a model in which BI responses cluster according to the six specific contexts, namely, unfamiliar adults, approaching peers, performing in front of others, separation and preschool situations, unfamiliar situations in general, and novel physical activities suggestive of minor risk. It should be noted that although the overall chi-square value indicated a significant difference between the parameters of the data and the model, Marsh et al. (1988) emphasized the difficulty in obtaining nonsignificant chi-square values with large sample sizes. Thus, in view of the fit indexes, it would be inappropriate to reject Model 3 on the basis of the chi-square statistic. When viewed as a whole, the fit indexes indicated that the six-correlated-factor model provided a relatively good fit of the mother, father, and teacher data. The six latent factors were all significantly correlated with each other ($p < .001$). It is not surprising that for all informants, the strongest correlations were between the factor assessing BI in general novel situations and those assessing BI in specific situations relating to unfamiliar adults, separation and preschool, and performance in front of others (see Table 3). The weakest correlations were between the physical challenge factor and factors relating to BI in separation, performance, and unfamiliar adult situations.

Model 3 (six correlated factors loading onto one higher order factor). The higher order model examined the degree to which the intercorrelation among the six factors could be explained by a single second-order factor representing a general dimension of BI. Table 1 shows values for NNFI, CFI, and Satorra-Bentler fit indexes close to .90, indicating that the second-order factor model also explained the data well. However, it is important to note that the fit of the second-order model cannot exceed the fit of the corresponding model involving the correlated first-order factors alone. Higher order factors are merely attempting to explain the covariation between the first-order factors (Marsh & Hocevar, 1985). Thus, it is not appropriate to compare the level of fit of the second-order model versus the first-order model. Rather, the aim is to determine whether the higher order model provides a satisfactory explanation for the covariance between the first-order factors. To do so, Marsh and Hocevar (1985) suggested the use of a target coefficient, that is, the ratio of the chi-square value of the first-order model to the chi-square of the second-order model. The target coefficient has an upper limit of 1, which would be possible only if the covariance between first-order factors could be

Table 3
Standardized Intercorrelations Among the Six Latent Factors and Loadings on Higher Order Factor

	Mother						Father						Teacher					
	Peers	Physical	Sep/pre	Perform	Adults	General	Peers	Physical	Sep/pre	Perform	Adults	General	Peers	Physical	Sep/pre	Perform	Adults	General
Peers	1.00						1.00						1.00					
Physical	0.40	1.00					0.52	1.00					0.76	1.00				
Sep/pre	0.62	0.26	1.00				0.66	0.43	1.00				0.75	0.59	1.00			
Perform	0.66	0.28	0.50	1.00			0.62	0.30	0.42	1.00			0.89	0.66	0.70	1.00		
Adults	0.71	0.30	0.50	0.62	1.00		0.67	0.37	0.45	0.54	1.00		0.90	0.64	0.69	0.87	1.00	
General	0.90	0.52	0.76	0.71	0.79	1.00	0.89	0.65	0.79	0.65	0.74	1.00	0.93	0.81	0.89	0.85	0.84	1.00
Loading on higher order	0.90	0.48	0.73	0.72	0.79	1.00	0.89	0.61	0.76	0.65	0.73	1.00	0.97	0.78	0.82	0.90	0.90	0.97
% unique variance	19	77	47	48	38	1	21	63	42	58	47	1	6	39	33	19	19	6

Note. All correlations between factors and factor loadings significant at $p < .001$. Sep/pre = separation/preschool.

Table 4
Cronbach's Alpha and Item-Total Correlations for Six Factors

	Mother ($n = 611$)		Father ($n = 506$)		Teacher ($n = 585$)	
Factor	Alpha	Item total	Alpha	Item total	Alpha	Item total
Total behavioral inhibition	.95	.30 – .81	.94	.31 – .75	.97	.52 – .83
Peer situations	.90	.66 – .78	.88	.65 – .76	.91	.70 – .79
Physical challenge	.80	.55 – .73	.72	.41 – .61	.86	.66 – .78
Separation/preschool	.90	.75 – .82	.86	.70 – .73	.89	.72 – .79
Unfamiliar adults	.91	.74 – .82	.89	.69 – .78	.95	.85 – .89
General novel situations	.90	.50 – .81	.88	.45 – .76	.91	.58 – .82

totally explained by the second-order factor. A target coefficient greater than .90 suggests that the second-order factor provides a good explanation for the covariance between factors. Comparison of the chi-square values of the six-correlated-factor model and the higher order model produced a target coefficient of .96 for both mother and father report. This result suggested that the higher order model did indeed provide a satisfactory explanation for the covariance between first-order factors and an adequate fit of the data for all mothers and fathers. For teachers, the target coefficient was slightly lower, but the overall fit of the higher order model suggested that it provided a good fit of the data for teacher report.

For all informants, the loadings of all first-order factors on the second-order factor were significant ($p < .01$; see Table 3). The extremely high loading between the novel situations in the general factor and the higher order factor, with almost zero unique variance remaining to be explained by the first-order factor, suggests that this first-order factor is assessing the same dimension as the second-order factor. In contrast, the percentages of unique variance accounted for by the other five factors, for mother and father report, were sufficient to justify their separate consideration, albeit as strong indicators of the overall construct of BI. For teacher report, the percentages of unique variance explained by the first-order factors were lower than that found for mother and father report, as shown in Table 3. This may reflect the more limited situations in which the preschool teachers observed the child. Taken together, the results support the use of a total BI score for all informants, made up of the sum of BI scores from the six contextual factors.

Internal Consistency

The internal consistency of the total score and each of the six factors of the BIQ was calculated. Table 4 outlines Cronbach's alpha and item-total correlations for each of the six factors. According to criteria used by Bickman et al. (1998), the alpha coefficients indicate acceptable internal consistency ($> .80$) for the total BIQ score and all factor scores for all informants, with the exception of father report on the physical activity inhibition factor, which was barely acceptable (.72).

Stability Over 12 Months

The stability coefficients of subscale and total BIQ scores over 12 months ranged from .49 to .78, depending on informant, as shown in Table 5. These values indicate moderate stability of BI over the 12-month retest period. It should be noted that for teachers it was not always possible to ensure that the same person completed the follow-up assessment, given staff turnover, and this is likely to explain the lower stability coefficients for teacher report.

Relationships Among Mother, Father, and Teacher Forms of BIQ

The mother and father total scores on BIQ were significantly correlated ($r = .69$, $p < .001$, $n = 499$; ICC =

Table 5
Twelve-Month Stability for Behavioral Inhibition Questionnaire (BIQ) Total and Factor Scores

	Mother ($n = 481$)	Father ($n = 381$)	Teacher ($n = 432$)
Total BIQ	.78***	.74***	.58**
Peer situations	.71***	.66***	.52***
Physical challenge	.65***	.55***	.50***
Separation/preschool	.60***	.64***	.42***
Unfamiliar adults	.74***	.65***	.57***
General novel situations	.75***	.68***	.49***

**$p < .01$ (two-tailed). ***$p < .001$ (two-tailed).

Table 6
Means and Standard Deviations for Mother, Father, and Teacher Ratings of Behavioral Inhibition for Total Sample and Behavioral Observation Group by Gender

		Mother				Father				Teacher			
		Total sample ($N = 611$)		Observed sample ($n = 100$)		Total sample ($N = 506$)		Observed sample ($n = 85$)		Total sample ($N = 585$)		Observed sample ($n = 98$)	
Factor		*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Peer situations	Boys	18.99	8.30	18.32	8.32	19.04	7.33	17.91	7.70	22.84	8.57	20.96	8.62
	Girls	19.79	8.12	21.04	7.77	19.41	7.33	21.45	7.34	23.49	9.52	24.85	8.58
	Total	19.40	8.21	19.60	8.14	19.23	7.32	19.49	7.70	23.17	9.06	22.79	8.78
Physical challenge	Boys	9.11	5.08	9.18	4.74	9.65	4.09	9.02	3.58	11.09	5.46	13.00	3.34
	Girls	9.70	4.73	9.91	4.71	10.39	4.43	9.87	3.86	12.52	5.90	14.22	3.75
	Total	9.41	4.91	9.52	4.72	10.02	4.28	9.40	3.71	11.82	7.73	13.57	3.57
Separation/preschool	Boys	11.97	6.37	10.16	5.98	11.54	5.23	10.68	5.45	13.00	5.95	11.88	7.03
	Girls	11.84	6.43	12.69	6.38	11.85	5.52	12.39	5.78	12.98	6.25	14.51	6.43
	Total	11.90	6.39	11.35	6.27	11.70	5.38	11.45	5.63	13.00	6.09	13.12	6.85
Performance	Boys	13.06	5.31	12.11	5.38	12.79	4.59	12.21	5.08	15.32	6.74	12.92	6.40
	Girls	12.15	5.36	13.81	6.06	12.04	5.02	13.45	5.79	15.28	7.15	15.72	6.88
	Total	12.60	5.35	12.91	5.74	12.41	4.82	12.76	5.41	15.30	6.94	14.24	6.74
Unfamiliar adults	Boys	13.59	6.09	12.81	6.09	14.29	5.56	13.77	6.38	16.16	7.00	14.12	7.13
	Girls	14.68	6.64	16.55	7.15	15.39	5.86	16.99	6.08	16.40	7.50	17.83	7.88
	Total	14.15	6.40	14.57	6.84	14.85	5.73	15.21	6.42	16.28	7.25	15.86	7.68
General novel situations	Boys	23.00	8.77	22.73	8.62	23.97	7.60	23.43	8.52	20.17	8.01	18.10	8.75
	Girls	23.79	9.25	26.88	8.95	24.24	8.33	26.24	8.97	19.96	8.82	21.55	9.26
	Total	23.40	9.02	24.68	8.98	24.11	7.97	24.68	8.78	20.06	8.43	19.72	9.11
Total behavioral inhibition	Boys	89.71	31.32	85.31	29.70	91.27	26.68	87.02	27.72	98.83	35.84	90.98	35.73
	Girls	91.96	32.64	100.90	31.61	93.31	29.12	100.38	27.69	100.41	40.39	108.67	36.67
	Total	90.85	31.99	92.64	31.45	92.31	27.94	93.00	28.34	99.63	38.19	99.28	37.07

.68, 95% CI = .63–.72), as were mother and teacher forms ($r = .47$, $p < .001$, $n = 577$; ICC = .46, 95% CI = .40–.53) and father and teacher forms ($r = .43$, $p < .001$, $n = 480$; ICC = .41, 95% CI = .33–.48).

Gender Effects

Table 6 shows the mean scores for subscale and total BIQ scores. Girls were rated by mothers as significantly more inhibited in performance, $F(1, 611) = 4.45$, $p < .05$, and adult situations, $F(1, 611) = 4.47$, $p < .05$, but not on any of the other subscales or the total BIQ score. For father report, girls were rated significantly higher than boys only on BI in adult situations, $F(1, 506) = 4.69$, $p < .05$. Teachers rated girls as significantly more inhibited only in physical situations, $F(1, 585) = 16.81$, $p < .001$. The gender effects evident in the reports from all informants were small in absolute terms, despite being statistically significant.

Convergent Validity

To assess its convergent validity, the parent and teacher forms of the BIQ were compared with the inhibition subscale of the TABC–R (Presley & Martin, 1994). The correlations among these measures were .87 for mother report, .86 for father report, and .85 for teacher report, indicating strong convergent validity, $p < .001$ in all cases.

Correlations Between BIQ and Behavioral Observation Measures

The mean BIQ scores for the sample of children who participated in the behavioral observation are shown in Table 6. There were no significant differences between the observation sample and children who were not observed on any of the subscales or total score of the BIQ, indicating that in terms of BI the observed preschoolers were representative of the sample in general. For the full

Table 7
Correlations Between Behavioral Observation Measures and Total Behavioral Inhibition Questionnaire (BIQ) Scores for Full Observation Sample (N = 100) and High- and Low-Behavioral-Inhibition (BI) Children

	Mother		Father		Teacher	
Behavioral measure	Full sample ($N = 100$)	High- and low-BI only ($n = 43$)	Full sample ($N = 85$)	High- and low-BI only ($n = 34$)	Full sample ($N = 98$)	High- and low-BI only ($n = 37$)
Contact initiation latency	.33**	.47**	.11	.15	.34**	.53***
Observer rating of BI	.46**	.60***	.25*	.35*	.28**	.47**
Number of prompts used	.33**	.55***	.15	.21	.27**	.44**
Speech						
Frequency of speech	−.30**	−.40**	.05	.10	−.16	−.27
Total duration of speech	−.33**	−.48***	−.13	−.19	−.25*	−.31
Mean duration of speech	−.36**	−.52***	−.28**	−.42*	−.32**	−.40*
Percentage of speech	−.33**	−.48***	−.13	−.19	−.25*	−.31
Eye gaze						
Frequency of eye gaze	−.02	−.07	.02	.01	.04	.10
Total duration of eye gaze	.19	.28	.21	.41*	.15	.20
Mean duration of eye gaze	.15	.20	.18	.43*	.04	.04
Percentage of eye gaze	.19	.28	.21	.41*	.15	.20

*$p < .05$ (two-tailed). **$p < .01$ (two-tailed). ***$p < .001$ (two-tailed).

sample of 100 observed children, significant correlations were found between BIQ ratings and several observational measures of child behavior, as shown in Table 7. BIQ ratings from all three informants were significantly correlated with the ratings of BI made by the independent observer from the videotaped interaction. In terms of specific behaviors, children who were rated by their mothers or teachers as demonstrating high BI took significantly longer to initiate verbal contact with the stranger, spoke for shorter durations of utterances (and therefore for less time in total), and required more prompting by the stranger to elicit speech. However, for father report, mean duration of speech episodes was the only behavior to correlate significantly with BIQ ratings for the total sample. Inspection of the scatterplots suggested that the significant correlations between observed behaviors and parent and teacher ratings of BI were mainly driven by the data points for children who showed extreme levels of BI (such as extremely low mean speech duration or high number of prompts). The correlation between ratings of BI and behavioral indicators was much weaker for participants whose behavior was in the midrange. To illustrate this point, the scatterplots for mother and teacher ratings of BI in relation to number of prompts and mean duration of speech are shown in Figure 1.

Given that some researchers prefer to regard BI as a categorical rather than continuous variable, the data set was divided into high and low BI using the top and bottom 20% of scores on the BIQ as the cutoff for group membership. The correlations were then examined for each of the behavioral measures with ratings on the BIQ, using the combined high- and low-BI sample. Table 7 indicates that, for the combined high- and low-BI groups, the correlations between BIQ scores and behavioral indicators of BI were higher, although consistent with those reported for the full observation sample. It was also interesting to find that fathers were more sensitive to their child's BI status when the extreme groups were considered, with father ratings of BI being significantly correlated with length of eye gaze episodes and total duration of eye contact, in addition to mean duration of speech and independent observer ratings of BI.

Relationships Between Categorical Classification of BI From Parent Questionnaires and Behavioral Observations

Mean values for each behavioral measure were compared across high- and low-BI groups, as shown in Table 8. According to mother BIQ categories, inhibited children as a group took significantly longer to initiate interaction, were rated as more behaviorally inhibited by the observer, required more prompts during the interaction, spoke less frequently and for shorter utterances, and showed longer eye gaze. The pattern of results was similar

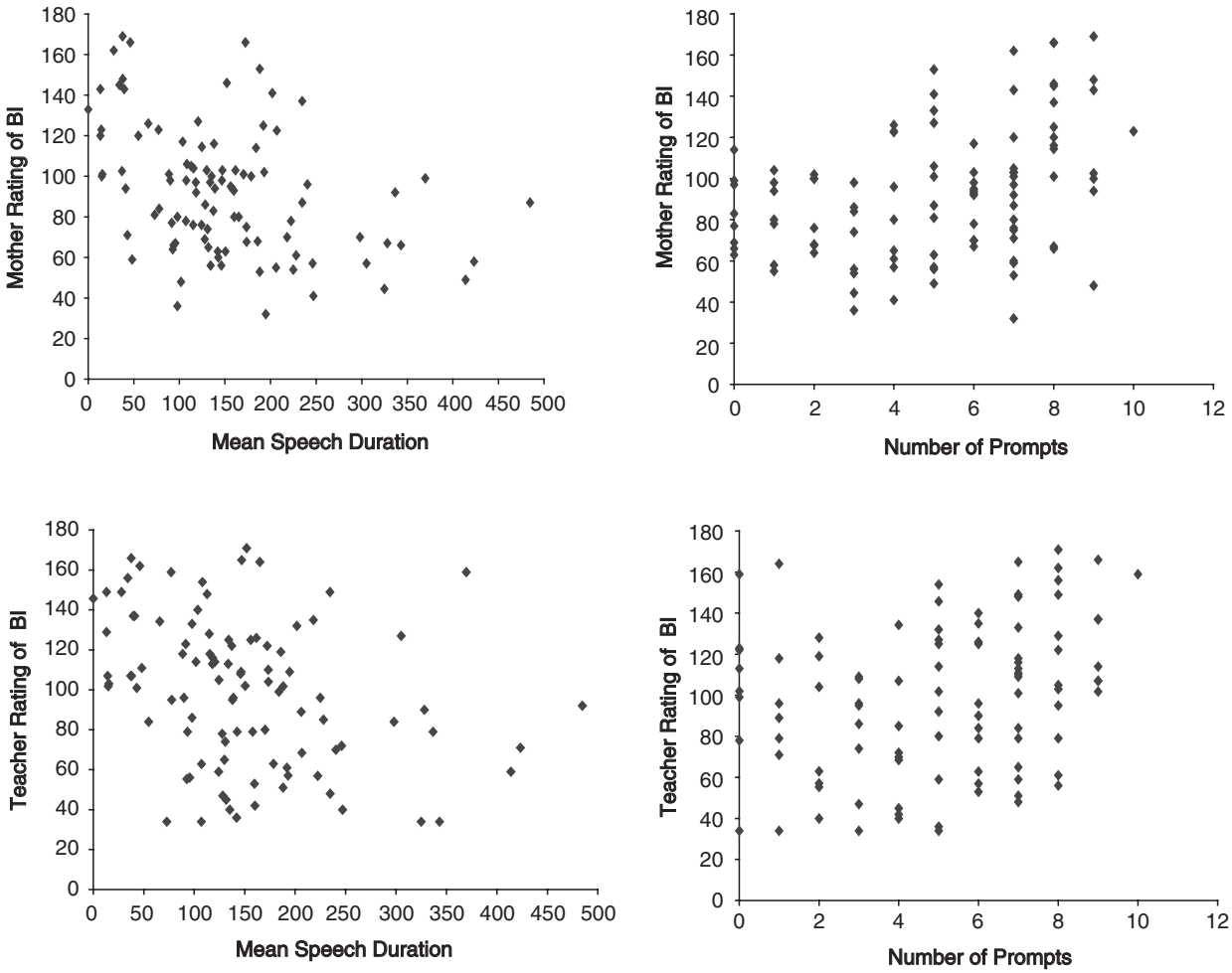

Fig. 1. Scatterplots of mother and teacher ratings of behavioral inhibition as a function of number of prompts and mean duration of speech episodes during stranger–child interaction.

when teacher BIQ was used to categorize high- and low-BI status, although differences in eye gaze were not evident. When father report on BIQ was used to identify high- and low-BI status, the groups differed in terms of the observer rating of BI, mean duration of speech utterances, and eye gaze duration.

Discussion

The aim of this study was to determine whether parents and teachers could provide a reliable and valid indicator of the temperament characteristic of BI in children. The first step of the study involved the development of a BI questionnaire that was economical, simple, and quick to administer and score. The items were selected from the research literature indicating features of BI. After piloting, the final version of the BIQ consisted of 30 items (28 items for the teacher version) that were selected to assess children's behavior across six contexts reflecting three domains in which inhibition has been found to occur.

The BIQ was found to have relatively robust psychometric properties. Confirmatory factor analysis supported a model based on six correlated factors reflecting the six contexts examined. For mothers, fathers, and teachers, items loaded strongly and significantly on factors reflecting the various inhibition contexts that the items purported to measure, namely, situations involving unfamiliar adults, peers, performing in front of others, separation and preschool, novel situations in general, and physical activities suggestive of minor risk. Further confirmatory factor analysis suggested that the intercorrelation between factors could be satisfactorily explained by a single, second-order factor representing a general dimension of BI. Each first-order factor loaded significantly on the higher order factor,

Table 8
Behavioral Observation Measures Across High- and Low-Behavioral-Inhibition (BI) Groups

	Mother BIQ			Father BIQ			Teacher BIQ		
Observed behaviour	High BI ($n = 22$) M (SD)	Low BI ($n = 21$) M (SD)	$F(1, 41)$	High BI ($n = 15$) M (SD)	Low BI ($n = 17$) M (SD)	$F(1, 30)$	High BI ($n = 17$) M (SD)	Low BI ($n = 21$) M (SD)	$F(1, 36)$
Contact initiation latency	149.45 (59.15)	77.14 (81.39)	11.18**	110.60 (70.03)	71.76 (73.41)	2.33	158.94 (55.72)	767.53 (83.70)	13.87***
Observer rating of BI	5.14 (1.52)	2.43 (1.43)	35.99***	4.27 (2.15)	2.65 (1.58)	5.98*	5.00 (1.75)	3.10 (1.45)	12.28***
No. of prompts used	7.00 (1.85)	4.33 (2.13)	19.26***	6.13 (2.92)	4.59 (2.12)	2.98	7.13 (2.16)	4.68 (2.16)	11.10***
Speech (SP)									
SP frequency	11.09 (5.94)	16.62 (6.16)	8.98**	14.40 (8.44)	14.29 (4.82)	.002	12.00 (6.84)	15.68 (5.87)	2.94
Total SP duration (ms)	1279.18 (1429.02)	3591.43 (2413.28)	14.78***	1767.20 (1458.54)	2474.71 (1584.27)	1.71	1211.81 (1292.94)	2709.57 (1337.62)	11.25**
Mean SP duration (ms)	94.02 (77.20)	210.13 (102.89)	17.63***	102.57 (70.45)	171.26 (106.09)	4.52*	83.77 (65.74)	177.16 (84.69)	12.89***
Percentage of SP	71.07 (79.39)	199.52 (134.07)	14.78***	98.18 (81.03)	137.48 (88.02)	1.71	67.32 (71.83)	150.53 (74.31)	11.25**
Eye gaze (EC)									
EC frequency	14.82 (6.52)	16.00 (7.29)	0.31	14.20 (5.44)	14.06 (4.46)	0.01	15.12 (6.04)	15.16 (5.77)	0.31
Total EC duration (ms)	9395.23 (4844.85)	6003.71 (4295.83)	5.88*	9776.40 (4485.14)	6253.71 (4235.15)	5.22*	8978.25 (4788.79)	7235.63 (3987.30)	1.38
Mean EC duration (ms)	712.41 (499.84)	393.57 (304.77)	6.30*	736.58 (411.83)	431.89 (245.55)	6.65*	603.28 (354.91)	531.54 (471.29)	0.25
Percentage of EC	521.96 (269.16)	333.54 (238.66)	5.88*	543.13 (249.17)	347.43 (235.29)	5.22*	498.79 (266.04)	401.97 (221.52)	1.38

Note. BIQ = Behavioral Inhibition Questionnaire.
*$p < .05$. **$p < .01$. ***$p < .001$.

supporting the use of a total BI score. However, five of the six factors retained sufficient unique variance to suggest that rather than being regarded as a single dimension, BI should be regarded as a construct in which there is a degree of situational specificity of behavior that is strongly influenced by an overall tendency toward inhibited behavior. This finding is in keeping with the proposition of Kagan et al. (1998) that inhibited children can display an avoidant style in several different contexts but not necessarily in all of them. These authors suggested that although there is a general tendency for the inhibited child to react in an avoidant manner to unfamiliar people, places, and situations, a child may, with experience, learn to control inhibition in some contexts but not others. Thus, although there is a strong overriding higher order factor of BI in general, a degree of situational specificity of behavior across different types of novel situations is also evident, justifying both the total and subscale scores.

The internal consistency of total BI score and subscale scores was acceptable for all informants. The stability of the BIQ over 12 months was also acceptable for mother and father report. For the teacher BIQ, stability of BIQ scores was lower, but this was to be expected given that the children's teachers differed across the two occasions. It is also possible that the higher stability of the parent ratings reflects a greater tendency for parents to form a stereotyped view of the child that is resistant to change over time. The present data do not allow us to exclude this possibility. The agreement among mother, father, and teacher report on the BIQ was relatively good. In terms of concurrent validity, the BIQ correlated strongly and significantly with the inhibition subscale from a child temperament questionnaire.

Having established the acceptable psychometric properties of the BIQ, the study then examined the construct validity of parent and teacher reports of BI by evaluating the extent to which parent and teacher report was sensitive to the observed behaviors typical of BI in children. Significant, positive correlations were found between the questionnaire ratings of all informants and the independent observer ratings of BI from the videotaped interaction. In terms of behavioral measures, children who were rated by their mothers and teachers as highly behaviorally inhibited took longer to initiate contact with the stranger, spoke less often and for shorter periods (and for less time in total), and required more prompting by the stranger to elicit speech compared with children who were rated as low on the BIQ. This result is consistent with the findings of Asendorpf (1993b) and Kagan et al. (1988) in comparing BI and non-BI children. In addition, children rated as high BI by mothers and fathers were found to show significantly longer eye gaze compared with low-BI peers, in keeping with the results of Gersten (1989) and Hill et al. (1999).

In general, fathers appeared to be less sensitive to their child's BI features, with father report showing lower association with observational indicators of their child's behavior and a lower level of agreement with teacher ratings as compared with mother report. It is possible that fathers less frequently observe their children in situations that are likely to trigger BI behavior than do mothers or teachers. It also appears that fathers may focus on a different set of behaviors in formulating judgments about their child's BI, being more sensitive to nonverbal features such as eye gaze rather than verbal aspects such as initiation latency, number of prompts required, and frequency of speech. Although this hypothesis is speculative, it would be worth exploring in future research.

The findings were consistent, regardless of whether BI was analyzed as continuous or as a categorical variable, although the correlations were stronger when the sample was restricted to children at the extremes of inhibition. Inspection of the scatterplots suggested that the significant correlations between observed behaviors and parent and teacher ratings of BI were mainly driven by the data points for children who showed extreme levels of inhibited behavior (such as extremely low mean speech duration or high number of prompts). It may be that parents and teachers attend less to the overt features of BI in the midrange, being more sensitive to behaviors at the extremes. For example, adults may be more discriminating among high or low levels of speech or social interaction in children, with midrange values being relatively unnoticed. Alternatively, behavior across situations may be more consistent for children who show very high or low levels of BI in the stranger–child situation compared with children who showed midrange behavior. Parent and teacher ratings of BI are likely to be based on their perceptions of child behavior across a broad range of situations and thus would correlate more highly with observed behavior of children who showed extremely high or low BI in the observation task. These possible interpretations of the findings are speculative and warrant further examination in future research.

The design of the present study does not permit us to draw conclusions about the question of whether BI is best regarded as a categorical construct rather than a continuous dimension of a human

temperament characteristic. To address this question, one would need to determine whether the concomitants of BI, such as physiological parameters, are evident only for children with an extreme level of BI, with minimal relationship with BI scores for children outside of the extreme range (see Kagan et al., 1994, for discussion of this issue). The present study does not present a test of the continuous versus categorical debate because both indicators are behavioral, with only the method of assessment being varied (questionnaire vs. observation).

It should be noted that although many of the correlations between parental or teacher report on the BIQ and observed behavior were statistically significant, a good deal of variance in observed behavior could not be explained by questionnaire reports. To some extent this is not surprising given that parent and teacher report of BI integrates observations of the child from a wide range of naturalistic settings. In contrast, the behavioral observation task was limited to one context, and although the task itself reflected a novel situation, we cannot be sure that the child's behavior was representative of his or her usual behavior across a range of novel settings.

The study does, of course, have strengths and weaknesses. Its positive features include the large sample size, acceptable psychometric properties of measures, and high interrater reliability of observations. The inclusion of father report and multiple informants is a further strength. Its limitations include the restricted age range of the children and a relatively narrow spectrum of socioeconomic status and ethnicity. It remains to be shown whether the findings would generalize to other age groups and sociodemographic characteristics. Furthermore, the observations were limited to a simulated interaction with an adult stranger, whereas the questionnaire assessment of BI included a broader spectrum of situations relating to peers, performance, physical challenge, and separation and preschool. For a more rigorous test of construct validity, it would have been preferable for the observational tasks to more closely reflect the situations covered in the BIQ. In addition, the observation period was brief and much shorter in duration than the multiple-task laboratory observations used to assess BI by Kagan and colleagues (Kagan et al., 1984; Kagan et al., 1988). Unfortunately, the resources of the project did not permit longer and more detailed observations.

An issue that needs to be addressed in future research is whether parent report of their child's BI is as effective as behavioral observation methods in accurately predicting which children are at risk for adverse outcomes. For example, Schwartz et al. (1999) reported a significantly higher prevalence of anxiety disorders among adolescents who had been categorized as high BI from behavioral observation during infancy compared with those who had been categorized as low BI. Similar studies are required that determine whether parental report of BI is as accurate as measurements of BI based on behavioral observation in predicting who will develop emotional and behavioral problems in the longer term. Further research is also required to determine whether parent and teacher questionnaire report can provide a reliable and valid indictor of BI among children from other age groups and cultural backgrounds.

In conclusion, results of the present study support the proposition that parents and teachers can indeed provide a reliable and valid report of children's BI. This finding is encouraging in that parent and teacher report may provide a relatively economical and simple method of assessing childhood BI. If further research can substantiate these results, parent and teacher report may provide a more cost-effective alternative to expensive and complex observational methods for identifying BI children. As such, it would provide a more feasible screening method that could be used within preschool contexts to identify children at risk for a range of adverse outcomes as a result of a high-BI temperament during early childhood.

References

Asendorpf, J. B. (1987). LOGIC—Report No. 2: Assessment procedures used in waves one to three. In F. E. Weinert & W. Schneider (Eds.), *Social competence* (pp. 53–125). Munich, Germany: Max Planck Institut fur Psychologische Forschung.

Asendorpf, J. B. (1990a). Beyond social withdrawal: Shyness, unsociability, and peer avoidance. *Human Development*, *33*, 250–259.

Asendorpf, J. B. (1990b). Development of inhibition during childhood: Evidence for situational specificity and a two-factor model. *Developmental Psychology*, *26*, 721–730.

Asendorpf, J. B. (1992). Beyond stability: Predicting interindividual differences in intraindividual change. *European Journal of Personality*, *6*, 103–117.

Asendorpf, J. B. (1993a). Abnormal shyness in children. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *34*, 1069–1081.

Asendorpf, J. B. (1993b). Beyond temperament: A two-factorial coping model of the development of inhibition during childhood. In K. H. Rubin & J. B. Asendorpf

(Eds.), *Social withdrawal, inhibition, and shyness in childhood* (pp. 265–289). Hillsdale, NJ: Erlbaum.

Asendorpf, J. B. (1994). The malleability of behavior inhibition: A study of individual developmental functions. *Developmental Psychology, 30*, 912–919.

Australian Bureau of Statistics. (1997). *ASCO: Australian standard classification of occupations* (2nd ed.), Canberra, Australia: Author.

Ball, C. T., Pelco, L. E., Havill, V., & Reed, V. E. (2001). Confirmatory factor analysis of the temperament assessment battery for children- revised: Parent form. *Journal of Psychoeducational Assessment*, *19*, 365–379.

Bentler, P. M. (1985). *Theory and implementation of EQS: A structural equations program*. Los Angeles, CA: BMDP Statistical Software.

Bickman, L., Nurcombe, B., Townsend, C., Belle, M., Schut, J., & Karver, M. (1998). *Consumer measurement systems in child and adolescent mental health*. Canberra, Australia: Department of Health and Family Services.

Biederman, J., Rosenbaum, J., Bolduc-Murphy, E. A., Faraone, S. V., Chaloff, J., Hirshfeld, D. R., et al. (1993). A three year follow-up of children with and without behavioral inhibition. *Journal of the American Academy of Child and Adolescent Psychiatry, 32*, 814–821.

Block, J. (1978). *The Q-Sort method in personality assessment and psychiatric assessment*. Palo Alto, CA: Consulting Psychologists Press.

Block, J. H., & Block, J. (1980). The role of ego-control and ego-resiliency in the organisation of behavior. In W. A. Collins (Ed.), *Minnesota Symposium on Child Psychology, Vol. 13* (pp. 39–101). Hillsdale, NJ: Erlbaum.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 247–261). Newbury Park, CA: Sage.

Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications and programing*. Thousand Oaks, CA: Sage.

DiLalla, L. F., & Falligant, E. L. (1995). An environmental and behavioral genetic perspective on behavioral inhibition in toddlers. In L. F. DiLalla & S. M. Dollinger (Eds.), *The assessment of biological mechanisms across the lifespan* (pp. 91–119). Hillsdale, NJ: Erlbaum.

Finegan, J., Niccols, A., Zacher, J., & Hood, J. (1989). Factor structure of the Preschool Characteristics Questionnaire. *Infant Behavior and Development*, *12*, 221–227.

Garcia Coll, C., Kagan, J., & Reznick, J. S. (1984). Behavioral inhibition in young children. *Child Development*, *55*, 1005–1019.

Gerbing, D. W., & Anderson, J. C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 40–65). Newbury Park, CA: Sage.

Gersten, M. (1989). Behavioral inhibition in the classroom. In J. S. Reznik (Ed.), *Perspectives on behavioral inhibition. The John D. and Catherine T. MacArthur Foundation series on mental health and development* (pp. 71–91). Chicago: University of Chicago Press.

Hegvik, R. L., McDevitt, S. C., & Carey, W. B. (1982). The Middle Childhood Temperament Questionnaire. *Journal of Developmental and Behavioral Pediatrics*, *3*, 197–200.

Hill, S. Y., Lowers, L., Locke, J., Snidman, N., & Kagan, J. (1999). Behavioral inhibition in children from families at high risk for developing alcoholism. *Journal of the American Academy of Child and Adolescent Psychiatry, 38*, 410–420.

Hirshfeld, D. R., Rosenbaum, J., Biederman, J., Bolduc, E. A., Faraone, S. V., Snidman, N., et al. (1992). Stable behavioral inhibition and its association with anxiety disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 31*, 103–111.

Kagan, J. (1992). Temperamental contributions to emotion and social behavior. In M. S. Clark (Ed.), *Emotion and social behavior. Review of personality and social psychology, Vol. 14* (pp. 99–118). Thousand Oaks, CA: Sage.

Kagan, J., Reznick, J. S., Clarke, C., Snidman, N., & Garcia Coll, C. (1984). Behavioral inhibition to the unfamiliar. *Child Development*, *55*, 2212–2225.

Kagan, J., Reznick, J. S., & Snidman, N. (1987). Temperamental variation in response to the unfamiliar. In N. A. Krasnegor, M. B. Elliott, A. H. Myron, & W. P. Smotherman (Eds.), *Perinatal development: A psychobiological perspective* (pp. 421–440). New York: Harcourt Brace Jovanovich.

Kagan, J., Reznick, J. S., Snidman, N., Gibbons, J., & Johnson, M. O. (1988). Childhood derivatives of inhibition and lack of inhibition to the unfamiliar. *Child Development*, *59*, 1580–1589.

Kagan, J., Snidman, N., & Arcus, D. (1998). Childhood derivatives of high and low reactivity in infancy. *Child Development*, *69*, 1483–1493.

Kagan, J., Snidman, N., Arcus, D., & Reznick, J. S. (1994). *Galen's prophesy: Temperament in human nature*. New York: Basic Books.

Kagan, J., Snidman, N., Zentner, M., & Peterson, E. (1999). Infant temperament and anxious symptoms in school age children. *Development and Psychopathology, 11*, 209–224.

Kochanska, G. (1991). Patterns of inhibition to the unfamiliar in children of normal and affectively ill mothers. *Child Development*, *62*, 250–263.

Kochanska, G., & Radke-Yarrow, M. (1992). Inhibition in toddlerhood and the dynamics of the child's interaction with an unfamiliar peer at age five. *Child Development*, *63*, 325–335.

Maccoby, E. E., & Martin, J. A. (1983). Socialization in the context of the family: Parent-child interaction. In E. M. Hetherington (Ed.), *Handbook of child psychology, Vol. 4. Socialization, personality, and social development* (pp. 1–102). New York: Wiley.

Marsh, H. M. (1994). Using the National Longitudinal Study of 1988 to evaluate theoretical models of self-concept: The self-description questionnaire. *Journal of Educational Psychology, 86*, 439–456.

Marsh, H. M., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, *103*, 391–410.

Marsh, H. M., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher-order factor models and their invariance across groups. *Psychological Bulletin*, *97*, 562–582.

McDonald, R. P., & Marsh, H. M. (1990). Choosing a multivariate model: Noncentrality and goodness-of-fit. *Psychological Bulletin*, *107*, 247–255.

Presley, R., & Martin, R. P. (1994). Toward a structure of preschool temperament: Factor structure of the Temperament Assessment Battery for Children. *Journal of Personality, 62*, 415–448.

Prior, M., Smart, D., Sanson, A., & Oberklaid, F. (2000). Does shy-inhibited temperament in childhood lead to anxiety problems in adolescence? *Journal of the American Academy of Child and Adolescent Psychiatry, 39*, 461–468.

Reznick, J. S., Kagan, J., Snidman, N., Gersten, M., Baak, K., & Rosenberg, A. (1986). Inhibited and uninhibited children: A follow-up study. *Child Development*, *57*, 660–680.

Robins, L. N. (1963). The accuracy of parental recall of aspects of child development and child-rearing practices. *Journal of Abnormal and Social Psychology, 33*, 261–270.

Rowe, D. C., & Plomin, R. (1977). Temperament in early childhood. *Journal of Personality Assessment*, *41*, 150–156.

Rubin, K. H., Hastings, P. D., Stewart, S. L., Henderson, H. A., & Chen, X. (1997). The consistency and concomitants of inhibition: Some of the children, all of the time. *Child Development*, *68*, 467–483.

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structural analysis. In *Proceedings of the Business and Economic Statistics Section of the American Statistical Association* (pp. 308–313). Alexandria, VA: American Statistical Association.

Schwartz, C. E., Snidman, N., & Kagan, J. (1999). Adolescent social anxiety as an outcome of inhibited temperament in childhood. *Journal of the American Academy of Child and Adolescent Psychiatry, 38*, 1008–1015.

Snidman, N. (1989). Behavioral inhibition and sympathetic influence on the cardiovascular system. In J. S. Reznick (Ed.), *Perspectives on behavioral inhibition* (pp. 51–70). Chicago: University of Chicago Press.

Thomas, A., & Chess, S. (1977). *Temperament and development*. New York: Brunner/Mazel.

Tryon, W. (1998). Behavioral observation. In A. S. Bellack & M. Hersen (Eds.), *Behavioral assessment. A practical handbook* (4th ed., pp. 79–103). Boston: Allyn & Bacon.

Woodward, S. A., Lenzenweger, M. F., Kagan, J., Snidman, N., & Arcus, D. (2000). Taxonic structure of infant reactivity: Evidence from a taxometric perspective. *Psychological Science*, *11*, 296–301.