Development and Validation of a Q–Sort Procedure to Assess Temperament and Behavior in Preschool-Age Children

Maureen E. Buckley, Daniel N. Klein, C. Emily Durbin, Elizabeth P. Hayden, and Kirstin C. Moerk
State University of New York at Stony Brook

Few broad observational measures of preschool-age children’s temperament and behavior currently exist. Studied the Child Temperament and Behavior Q–Set (CTBQ–Set) as a naturalistic observation measure to tap the major domains of temperament and behavior in preschoolers. Pairs of observers rated the behavior of a community sample of preschoolers during 2 independent home visits using q-sort methodology. The CTBQ–Set scales displayed good levels of internal consistency and convergent and discriminant content saturation. The scales displayed good interrater reliability at each observation and moderate test–retest stability between observations. The construct validity of the CTBQ–Set scales was supported by correlations with mothers’ ratings of their children’s behavior using the Child Behavior Checklist for 2- and 3-year-olds (CBCL/2–3) and the Children’s Behavior Questionnaire. The CTBQ–Set shows promise as an observer-rated measure of preschoolers’ behavior and temperament in their natural home environment.

A wide array of measures has been developed to assess the behavior, temperament, and adjustment of school-age children. In contrast, there is a dearth of such measures for preschool-age children. Behavior and adjustment problems in preschoolers have been demonstrated to be stable and clinically significant (e.g., Lavigne et al., 1998). Furthermore, the relation of temperament and behavior in the preschool period to personality and psychopathology in adolescence and young adulthood has been established (e.g., Caspi, Moffitt, Newman, & Silva, 1996; Caspi & Silva, 1995). Thus, there is a need to broaden the armamentarium of assessment instruments available for children in the preschool age range (Ablow et al., 1999).

Parental report questionnaires are the most widely used instruments for assessing behavior and temperament in preschoolers, as they offer a time- and cost-effective method of tapping extensive knowledge of the child’s behavior in diverse situations. The best-validated parental report measure of child behavior is the Child Behavior Checklist (CBCL; Achenbach, 1995), which assesses a broad array of internalizing and externalizing problems. The CBCL is available in a format specific to 2- and 3-year-old children (CBCL/2–3) and is useful for measuring developmentally appropriate behaviors over time, as it maps onto other forms of the CBCL for children up to age 16 (Achenbach, Edelbrock, & Howell, 1987).

However, there are limitations to using parent questionnaires as the sole approach to assessing preschoolers’ behavior (Kagan, 1998). Parental reports may be influenced by their level of education, social class, and personality characteristics or psychopathology. Additionally, parents may not be aware of what is and is not normative behavior at particular ages, and their interpretation of items may differ from what was intended by the developers of the instrument. These limitations, along with the historically poor agreement between direct observer and parent ratings of preschoolers’ behaviors (Kagan, 1998), suggest that the validity of parental report questionnaires may be limited. Thus, Kagan suggested that future research on preschoolers’ behaviors should employ direct observation measures, rather than relying solely on parental report.

Direct observation measures of behavior in preschoolers have received less attention than parental report questionnaires. However, laboratory assessment batteries, such as Goldsmith and Rothbart’s (1988) Laboratory Temperament Assessment Battery, are currently being developed to provide detailed assessments of temperamental constructs for infants and preschoolers.

The q-sort provides another approach to measuring preschoolers’ behavior. For example, the California Child Q–Set (CCQ) was developed to provide a means for professionals to summarize and integrate informa-
tion on the personality of children ages 3 to 11 obtained from diverse sources (Block & Block, 1969). The CCQ was modified to a “common language” version for use by parents or nonprofessional observers by Caspi and colleagues (1992). Although both versions of the CCQ have many uses, they were not developed to systematically sample the major domains of child temperament and behavior that have emerged in recent years (Rothbart & Bates, 1998; Shiner, 1998). Moreover, neither version was designed for use by research staff in a limited number of observation periods. A number of CCQ items refer to behaviors that require knowledge of the child over a longer time period than allowed by a single home visit (e.g., “has transient interpersonal relationships,” “most adults seem to like him”). Finally, as the CCQ covers a broad age range, a number of the items are not developmentally appropriate for most preschool-age children (e.g., “shows concern for moral issues,” “has a readiness to feel guilty”).

In light of the limited number of broadband observational measures of preschool-age children’s temperament and behavior, especially during a time-limited assessment period, we developed the Child Temperament and Behavior Questionnaire—Set (CTBQ–Set) as an observational instrument for use in naturalistic settings, such as the home.

The CTBQ–Set is a molar-level measure, in the sense that it is intended to be contextually sensitive and assesses a wide range of behaviors as they unfold over an extended period of observation (2 to 3 hr), as opposed to micro-level measures that assess the occurrence or frequency of a limited number of discrete behaviors during a series of brief time intervals (e.g., seconds or minutes). The CTBQ–Set covers five of the major dimensions of temperament and behavior that have recently been of greatest interest to investigators in child clinical psychology and personality development: positive affectivity (PA), negative affectivity (NA), sociability, separation anxiety/dependency, and externalizing behavior (Achenbach, 1995; Rothbart & Bates, 1998; Shiner, 1998).

Contemporary models of child temperament emphasize the centrality of individual differences in emotionality (Goldsmith, 1993; Rothbart & Bates, 1998), and most major child and adult temperament and personality systems include PA and NA (Clark & Watson, 1999; Rothbart & Bates, 1998). Within PA, separate subscales were included for mood, behavioral activation/impulsiveness, and appetitiv/approach behavior, as investigators have differed on which of these aspects are most central to the construct of PA (Depue & Collins, 1999; Lucas, Diener, Grob, Suh, & Shao, 2000; Russell & Carroll, 1999; Watson, Wiese, Vaidya, & Tellegen, 1999). NA includes separate subscales for internalizing (i.e., sadness or depression) and externalizing (i.e., anger or irritability) negative emotions, as these are distinguished in most temperament systems (McCrae & Costa, 1999; Shiner, 1998). We also included a subscale tapping deficits in the regulation of negative emotions (e.g., mood lability).

Sociability and separation anxiety/dependency overlap to some degree with PA and NA, but have a more interpersonal orientation. Some investigators have subsumed sociability under PA (Eysenck, 1991); however, more recent theorists have argued that although the constructs overlap, they have distinct correlates and implications (Depue & Collins, 1999; Lucas et al., 2000; Shiner, 1998). There is also evidence for a further distinction between low sociability per se and shyness (or social withdrawal with strangers; Asendorf, 1989; Harrist, Zaia, Bates, Dodge, & Petit, 1997; Shiner, 1998), which closely relates to Kagan’s construct of behavioral inhibition (Kagan, Reznick, & Snidman, 1988) and is currently of great interest to developmental psychologists and psychopathologists. Hence, separate subscales were included for low sociability and social withdrawal with strangers.

Separation anxiety/dependency are not included in most temperament models. However, these constructs have important implications for personality development and psychopathology, and disturbances in this area often lead to clinical referrals (Bornstein, 1992; Fischer, Himle, & Thyer, 1999). Finally, we included a dimension for externalizing behavior, aspects of which (e.g., activity, attention, impulsivity or constraint) have been included in many temperament models (e.g., Buss & Plomin, 1984; Goldsmith, 1996; Rothbart, Ahadi, Hershey, & Fisher, 2001), have considerable clinical significance, and are the focus of many parent- and teacher-report instruments of child behavior (Achenbach, 1995). Although the optimal approach to parsing this domain is controversial, we included subscales of hyperactivity–impulsivity, noncompliance, and aggression. Some empirical evidence exists for distinguishing between these overlapping constructs (Heubeck, 2000; Olson, Bates, Sandy, & Lanthier, 2000; Shiner, 1998), and they mirror the distinctions between attention-deficit-hyperactivity disorder, oppositional defiant disorder, and conduct disorder in the current classification system of childhood psychopathology (American Psychiatric Association, 1994).

The CTBQ–Set also includes several items that do not fall within these five domains but are potentially important to the development or expression of behavior (e.g., presence of a physical handicap). Several other behaviors of interest were excluded, as they are unlikely to be seen during a home observation (e.g., sleep disturbances, bedwetting).

CTBQ–Set items are rated on the basis of frequency, intensity, duration, and appropriateness of behaviors in the particular context in which they are observed. Although the measure requires some knowledge of the developmentally normative prevalence of behaviors to make fine-grained discriminations be-
tween the relative salience of those behaviors for the child, nonprofessional observers can readily use the CTBQ–Set.

Due to the q-sort methodology, the CTBQ–Set provides an ipsative, or person-centered, description of preschoolers’ behaviors, rather than a variable-centered approach. Each item is scored in the context of an item set with wide-ranging content to provide a comparison of qualities within, rather than between, preschoolers (Caspi et al., 1992; Waters & Deane, 1985).

CTBQ–Set items refer to specific and well-defined behaviors, and their standard language discourages idiosyncratic interpretation of the items (Waters & Deane, 1985; Westen, Mudderisoglu, Fowler, Shedler, & Koren, 1997). The items are rank-ordered into nine categories, ranging from those behaviors most salient or characteristic to least salient or characteristic of a particular preschool child. This produces a fixed distribution of rater judgments that reduces rater bias (Caspi et al., 1992; Waters & Deane, 1985); as there are a limited number of available slots in the extreme categories (most and least salient items), the tendency to rate behaviors as extreme is countered by the fixed distribution of each sorter’s ratings (Ozer, 1993).

The q-sort methodology has a number of advantages, including allowing for the subsequent analysis of variables that may not have been directly assessed during data collection, such as development of different scales years after data collection has been completed (Waters & Deane, 1985). Q-sort data may also be subjected to a variety of analytic strategies (e.g., item analysis, a priori scale construction, cluster analysis of individuals, factor analysis of items, and comparison to criterion sorts) not available when summarizing a wide array of information into a single score. Finally, the ipsative approach to measuring preschoolers’ behavior allows for the comparison and merging of different samples (Caspi et al., 1992; Waters & Deane, 1985).

It is important to note that no developmentally appropriate diagnostic system of psychopathology exists for the preschool age range, except for the Diagnostic Classification: 0–3 system, which was generated on the basis of uncontrolled clinical observations and currently has unknown reliability and validity (Mash & Dozois, 1996). Hence, the CTBQ–Set is not geared toward formal diagnoses. Rather, it assesses key dimensions that are likely to be of interest to clinical and personality researchers.

The purpose of this article is to describe the development of the CTBQ–Set as a measure of preschoolers’ behavior in a natural setting, examine its reliability, and explore its construct validity by examining its relation to maternal reports of their 3-year-olds’ behaviors as measured by the CBCL/2–3 and the Children’s Behavior Questionnaire (Rothbart et al., 2001).

**Method**

**Sample**

**Recruitment.** A community sample was employed for the scale development and initial validity analyses reported here, with the intention of later extending the results to clinical populations. Participants were 100 children (55 boys and 45 girls) between the ages of 2.5 and 4 years ($M = 3.6$ years, $SD = 0.3$), and their parents. We identified potential participants through several methods: by contacting local families via a commercial mailing list (51%) and through local newspapers advertisements and fliers in preschools, daycare centers, and the community (49%). Written informed consent was obtained from one of the child’s parents as part of a larger study of preschoolers’ behavior and temperament, and oral consent was obtained again prior to each home visit. All families participating in the study were paid for their time. We screened for the presence of gross cognitive impairment with the Peabody Picture Vocabulary Test (third edition; Dunn & Dunn, 1997). Children with significant medical problems, disabilities, or gross cognitive impairment were excluded.

**Demographics.** The children were mostly white (96%), and the sample’s mean Hollingshead (1975) score was 34.6 ($SD = 9.8$), corresponding to middle and working class (social class III). Most of the children (97%) came from two-parent families. The majority of mothers (58%) worked outside of the home to some extent; 15% worked more than 35 hr per week. Mothers’ average age was 32.5 years ($SD = 6.1$); fathers’ average age was 36.3 years ($SD = 5.5$). The sample’s demographics were generally representative of Suffolk County, New York.

**Assessment Procedures**

**Home observations.** Each child was observed during two 2- to 3-hr home visits. One family did not complete the second home visit. CBCL/2–3 and CBQ data were collected within 1 month of the first observation from 95 mothers.

Observer training included a 2-hr didactic introduction to the CTBQ–Set, observations of preschoolers through visits to local preschool classrooms and videotapes, and group meetings discussing ratings of specific children. Group training meetings focused on salient and nonsalient examples of each item, how to make discriminations between the relative salience of behaviors for a child, and practical difficulties more experienced raters had encountered while completing the CTBQ–Set for a child. Instructions for rating each item were written onto each card in the q-set (e.g., “emotionally constricted: child’s affect is characteristi-
cally indifferent, disengaged, or flat—note facial movements, gestures, vocalizations,” “acts to maintain social interaction: includes observer after becoming familiar—child engages others in conversation, asks and answers questions, makes directed comments”; see the Appendix). Prior to rating independently, each observer was required to observe a child and then complete the CTBQ–Sort along with an experienced rater. A general criterion for interrater agreement of at least .80 per observation was observed throughout the study.

Trained pairs of graduate and undergraduate students who had no access to the parent questionnaire data conducted the observations. A different pair of raters observed the children at the second visit, without knowledge of the first home-visit data. The two observations were generally scheduled for different times of the day and days of the week, to observe the child’s behavior in different contexts (e.g., a weekday and weekend observation). The mean time between the first and second home visits was 5.5 weeks ($SD = 4.5$ weeks). The average length of a home visit was 2 hr and 32 min ($SD = 22$ min). The observers watched the child during his or her normal daily routine, taking notes on behaviors germane to CTBQ–Set items every 5 min. Observers were instructed to remain relatively unobtrusive during the observation period (e.g., not playing with the child), so as not to bias their ratings or to elicit nontypical behaviors from the child.

Families were instructed to participate in their normal daily activities, whether staying at home, running errands, or playing with neighborhood children. Home observations typically took place in the home or yard of the family (80%); however, 4% took place exclusively outside of the home (e.g., at the park, pool, or store), and 16% took place both in and outside of the home setting. The children’s mothers were present at 98% of observations (although not always in the same room), whereas their fathers were only present at 25%. Older and younger siblings were present for at least part of 69% of observations. Other adults, such as neighbors or extended family members, and other children, such as friends or cousins, were present for at least part of home visits (32% and 26%, respectively).

After each observation, raters independently completed the CTBQ–Sort. Meetings were held after the completion of each sort to discuss rating discrepancies and prevent rater drift. Two independent sets of ratings were available from each home observation, yielding a total of four sets of ratings of each child’s behavior.

CTBQ-Sort. The CTBQ–Set consists of 90 items tapping the following domains: low PA (low appetitive behavior, behavioral impoverishment, and affective constriction); NA (sadness or depression, anger or irritability, and emotional regulation deficits); sociability (low sociability and fear of novel situations); separation anxiety/dependency; and externalizing behavior (hyperactivity–impulsivity, noncompliance, and aggression). For a complete listing of CTBQ–Set items, see the Appendix.

We used a rectangular sort distribution, with 10 items per category. After each observation, raters sorted the 90 CTBQ–Set items into three broad categories of behavior: those uncharacteristic, neither characteristic nor uncharacteristic, and those characteristic of the child. Each of the broad categories was then subdivided into three finer-grained categories of behavior.

“Behaviors not descriptive of the child” was divided into three categories: behaviors (1) extremely, (2) quite, or (3) fairly uncharacteristic or negatively salient of the child. The middle category was subdivided into behaviors (4) somewhat uncharacteristic or negatively salient; (5) relatively neutral or unimportant behaviors, items impossible to judge on the basis of observed information, or not consistently descriptive; or (6) somewhat characteristic of the child. “Behaviors characteristic of the child” was subdivided into behaviors (7) fairly, (8) quite, or (9) extremely characteristic or salient of the child. The finer-grained categories were rated such that items in the extreme categories (e.g., 1, 2, 8, 9) were more descriptive of the child than items placed in the middle categories.

After sorting the CTBQ–Set items, raters checked that each pile contained exactly 10 items. In the case of unequal distribution of items, raters were forced to make decisions about which behaviors were more or less descriptive of the child relative to other behaviors tapped by the CTBQ–Set. These decisions were made based on the frequency, intensity, and appropriateness of the behavior in a given context.

CBCL/2–3. The CBCL/2–3 (Achenbach et al., 1987) consists of 100 items measuring behavior and adjustment in children of this age range and is more developmentally appropriate for 3½-year-olds than the version of the CBCL for older children. Mothers rated each item on a scale from 0 (not true) to 2 (very or often true) for their child’s behavior in the past 6 months. The Internalizing Problems scale is comprised of the anxious/depressed and withdrawn subscales, and the Externalizing Problems scale consists of the aggressive and destructive behaviors subscales. Internal consistency in our sample was consistent with the literature (Externalizing Problems $\alpha = .87$; Internalizing Problems $\alpha = .85$).

CBQ. The CBQ consists of 194 items measuring a variety of behaviors and emotions in children ages 3 to 12 over the past 6 months (Rothbart et al., 2001). Mothers rated items on a scale from 1 (extremely untrue) to 7 (extremely true), indicating how closely the statement describes their child’s typical behavior. The CBQ includes 15 scales: Smiling/Laughter, Activity Level, Anger/Frustration, Positive Anticipation, Atten-
Statistical Analyses

The internal consistency of the CTBQ–Set scales was evaluated using coefficient alpha. Following Jackson (1970), we assessed convergent and discriminant content saturation between subscales and major scales (for example, the extent to which the low appetitive behavior, behavioral impoverishment, and affective constriction subscales correlate with the Low PA major scale, not including the subscale being examined, in comparison to their relation with other major scales) using Pearson’s *r*. Additionally, Pearson’s *r* was calculated for the associations between major scales and between subscales.

Interrater reliability was assessed using the intraclass correlation (ICC–Case 1) for each individual observation. We examined rank-order test–retest stability using Pearson’s *r* values, both attenuated and corrected for attenuation (i.e., rater unreliability). Mean-level stability between observations was assessed via paired *t* tests. For test–retest stability analyses, the average of both observers’ ratings at the first observation was compared to the average of both observers’ ratings at the second observation. The reliability of the CTBQ–Set scales using the averaged ratings from all available observers was measured using the ICC–Case 1 for multiple raters (Lahey, Downey, & Saal, 1983).

Construct validity was assessed by correlating each CTBQ–Set scale with maternal ratings of CBCL/2–3 Internalizing and Externalizing Problems and each CBQ scale. On the basis of the literature, we predicted that CBCL Internalizing Problems would correlate with CTBQ–Set Low PA, Sociability, and Separation Anxiety/Dependency, and that CBCL Externalizing Problems would correlate with CTBQ–Set Externalizing Behavior. We also hypothesized that CTBQ–Set NA would correlate with both CBCL scales due to the associations between CTBQ–Set sadness/depression and CBCL Internalizing Problems, and between CTBQ–Set angry/irritable and CBCL Externalizing Problems.

We also predicted a number of associations between the CTBQ–Set and the CBQ (Rothbart et al., 2001). CTBQ–Set Low PA was expected to correlate negatively with CBQ Smiling/Laughter, Positive Anticipation, and High and Low Intensity Pleasure. CTBQ–Set Sociability was hypothesized to correlate positively with CBQ Fear and Shyness. CTBQ–Set NA was predicted to correlate positively with CBQ Sadness, Fear, and Anger, and negatively with CBQ Soothability. CTBQ–Set Separation Anxiety/Dependency was expected to correlate with CBQ Shyness. Finally, CTBQ–Set Externalizing Behavior was hypothesized to correlate positively with CBQ Activity Level, Impulsivity, and Anger, and negatively with CBQ Attentional Focusing and Inhibitory Control.

Results

Scale Development

Scale construction. We followed Jackson’s sequential system of personality scale development (Jackson, 1970). Most items were generated by the authors; however, a few items were adapted from other sources. Initial item analyses were conducted on the basis of a priori scales corresponding to the major domains identified in the literature as described previously. Scales were revised three more times, on the basis of internal consistency, item-total correlations, and convergent and discriminant content saturation. As these data were used to revise the CTBQ–Set, they should be considered descriptive rather than tests of the measure’s validity.

Internal consistency. The median alpha for the five major CTBQ–Set scales was good (*α* = .91), ranging from moderate to excellent (*α* = .82 to .95). As would be expected, alphas were slightly lower for the CTBQ–Set subscales, which contain fewer items. However, the median alpha for the subscales was still moderate (*α* = .87), ranging from fair to excellent (*α* = .70 to .96). The lowest coefficient alphas were obtained from subscales containing only two to three items (see Table 1). Two items (oversensitivity and anxiety/fear) contribute to the internal consistency of the NA major scale but are not included in any of its subscales.

Intercorrelations of CTBQ–Set major scales. Correlations between major scales ranged from .01 to .58, with a median of .33 (*p* < .05; see Table 2). The strongest associations between major scales involved Low PA with Low Sociability, and Low Sociability with Separation Anxiety/Dependency.
Convergent content saturation. Each subscale was correlated with its major scale after removing that subscale’s items from the major scale. The median correlation between major scales and their respective subscales was .97, ranging from .87 to .99 (all \( p < .001 \)).

We also examined the intercorrelations of the subscales contained within each major scale. The median correlations between subscales belonging to the same major scale were .80 (NA subscales), .77 (Low PA subscales), .76 (Sociability subscales), and .54 (Externalizing Behavior subscales; all \( p < .001 \)). Separation Anxiety/Dependency has no subscales.

Discriminant content saturation. The first step in establishing discriminant content saturation was to correlate each CTBQ–Set major scale with the subscales from the other major scales. The median correlation between major scales and subscales from other major scales was .27 (\( p < .01 \)), ranging from .02 (Low PA with hyperactivity and noncompliance) to .58 (Externalizing Behavior with angry/irritable). Each subscale correlated more highly with its own major scale than with any of the other major scales (see Table 3).

Next, we examined the intercorrelations between the subscales from different major scales. The median Pearson’s \( r \) between subscales belonging to different major scales was .28 (\( p < .1 \)), ranging from .00 (behavioral impoverishment and noncompliance) to .69 (angry/irritable and noncompliance; see Table 3). Subscales correlated more highly with subscales from their own major scales than with subscales from other major scales, with one exception. The noncompliance subscale correlated more highly with the angry/irritable (.69) and emotional regulation deficits (.55) subscales of NA than with the hyperactivity (.46) and aggression (.54) subscales of Externalizing Behavior.

Interrater Reliability and Test–Retest Stability

Reliability of the CTBQ–Set scales was assessed by examining the interrater reliability for each home observation, the test–retest stability of ratings between observations, and the reliability of the composite ratings from all four observers (see Table 1).
observations, the best interrater agreement was found for the Low Sociability and Externalizing Behavior major scales and the fear of novel situations, low sociability, aggression, and noncompliance subscales.

**Test–retest stability.** The median uncorrected Pearson’s $r$ between home observations for the major scales was .43, ranging from .14 (Separation Anxiety/Dependency) to .51 (Sociability). For subscales, the median uncorrected Pearson’s $r$ was .40, ranging from .29 (behavioral impoverishment) to .51 (fear of novel situations). Correcting for rater unreliability, the median disattenuated Pearson’s $r$ between home observations for the major scales was .60, ranging from .24 (Separation Anxiety/Dependency) to .64 (Sociability). For subscales, the median disattenuated Pearson’s $r$ between home observations was .59, ranging from .43 (behavioral impoverishment) to .74 (affective constriction). Paired $t$ tests indicated that none of the major scales or subscales displayed a significant mean change between the first and second observations. We correlated the number of days between home visits with mean-level test–retest stability, as measured by the difference in CTBQ–Set scale ratings between the first and second observations. Only 1 of the 16 correlations was significant (emotional regulation deficits, $r = .22, p < .05$), which is no more than would be expected by chance.

**Reliability of four observer composite ratings.** Finally, we examined the scales’ reliability using the average of all available observer ratings. The median ICC for major scales was .81, (range = .66 to .86), and the median ICC for subscales was .77 (range =.71 to .86; see Table 1).

**Construct Validity**

**CBCL.** To assess construct validity, we examined the relation of the CTBQ–Set major scales and subscales to mothers’ ratings on the CBCL/2–3. Mean CBCL scale scores fell within the normal range for preschool-age children. Boys and girls did not significantly differ on mean Internalizing Problems (boys: $t = 46.0, SD = 7.7$; girls: $t = 47.8, SD = 11.4$) or Externalizing Problems (boys: $t = 46.2, SD = 7.6$; girls: $t = 47.1, SD = 8.9$) scale scores, and correlations between the CBCL and CTBQ–Set also did not differ significantly between the sexes. Therefore, boys’ and girls’ scores were combined for the following analyses (see Table 4).

The first set of relations we examined were between CTBQ–Set major scales and maternal CBCL ratings. As predicted, maternal CBCL ratings of Internalizing Problems correlated with CTBQ–Set Low PA and Low Sociability, and maternal CBCL ratings of Externalizing Problems correlated with CTBQ–Set Externalizing Behavior. CTBQ–Set NA was correlated with both CBCL dimensions. No other significant relations between the CTBQ–Set major scales and maternal CBCL scale ratings were obtained. However, several other significant relations emerged when comparing CTBQ–Set subscales with maternal CBCL ratings (see Table 4).

Maternal ratings for CBCL Internalizing and Externalizing Problems scales were highly correlated ($r = .55, p < .01$). Hence, we also computed partial correlations between the CTBQ–Set and each CBCL scale controlling for the other scale. These partial correlations allow us to examine the association between the CTBQ–Set and CBCL with greater specificity, as the nonspecific variance that is shared by both CBCL scales is partialed out.

To control for the influence of CBCL Externalizing Problems on the relation between the CTBQ–Set and CBCL Internalizing Problems, we examined the associations partialing out the influence of CBCL Externalizing Problems ratings. Examining partial $r$s, the significant relation between CBCL Internalizing Problems and CTBQ–Set Low Sociability was strengthened, and its relation with Low PA remained significant, but its relation with NA disappeared. A positive association between CBCL Internalizing Problems and CTBQ–Set Separation Anxiety/Dependency, and a

**Table 3. Median Correlations of Subscales With Their Own and Other Major and Subscales**

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Own Major Scale</th>
<th>Other Major Scales</th>
<th>Other Subscales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affective Constriction</td>
<td>.96</td>
<td>.31 (.06 – .54)</td>
<td>.32 (.01 – .55)</td>
</tr>
<tr>
<td>Behavioral Impoverishment</td>
<td>.97</td>
<td>.45 (.04 – .54)</td>
<td>.30 (.00 – .53)</td>
</tr>
<tr>
<td>Low Appetitive Behavior</td>
<td>.99</td>
<td>.44 (.10 – .50)</td>
<td>.25 (.02 – .49)</td>
</tr>
<tr>
<td>Sadness/Depression</td>
<td>.98</td>
<td>.37 (.26 – .53)</td>
<td>.37 (.11 – .50)</td>
</tr>
<tr>
<td>Angry/Irritable</td>
<td>.98</td>
<td>.21 (.10 – .31)</td>
<td>.31 (.03 – .69)</td>
</tr>
<tr>
<td>Emotional Regulation Deficits</td>
<td>.99</td>
<td>.26 (.10 – .51)</td>
<td>.31 (.05 – .55)</td>
</tr>
<tr>
<td>Low Sociability</td>
<td>.87</td>
<td>.38 (.26 – .57)</td>
<td>.24 (.10 – .55)</td>
</tr>
<tr>
<td>Fear of Novel Situations</td>
<td>.98</td>
<td>.37 (.20 – .48)</td>
<td>.22 (.03 – .46)</td>
</tr>
<tr>
<td>Hyperactivity</td>
<td>.93</td>
<td>.21 (.02 – .24)</td>
<td>.22 (.01 – .35)</td>
</tr>
<tr>
<td>Noncompliance</td>
<td>.95</td>
<td>.13 (.02 – .52)</td>
<td>.17 (.00 – .69)</td>
</tr>
<tr>
<td>Aggression</td>
<td>.97</td>
<td>.24 (.06 – .37)</td>
<td>.23 (.02 – .45)</td>
</tr>
</tbody>
</table>
negative association between CBCL Internalizing Problems and CTBQ–Set Externalizing Behavior also emerged. A similar pattern of results was evident when examining the partial correlations between CBCL Internalizing Problems and the CTBQ–Set subscales.

We also examined the correlations between the CTBQ–Set scales and the CBQ. There were no significant sex differences on any CBQ scales, or in the correlations between the CTBQ–Set and maternal CBQ ratings. Therefore, boys’ and girls’ scores were combined (see Table 5).

As predicted, CTBQ–Set Low PA was significantly negatively correlated with CBQ Smiling/Laughter but was not associated with CBQ Positive Anticipation, and High and Low Intensity Pleasure. CTBQ–Set NA was significantly associated with CBQ Sadness and Anger but not with Fear and Soothability. CTBQ–Set Sociability was significantly correlated with CBQ Shyness but not with Fear. CTBQ–Set Separation Anxiety/Dependency was significantly associated with CBQ Shyness. Finally, CTBQ–Set Externalizing Behavior was significantly positively correlated with CBQ Activity Level, Anger, and Impulsivity, and significantly negatively associated with Attentional Focusing and Inhibitory Control.

Several nonpredicted associations also emerged, although all seemed relatively consistent with the literature. CTBQ–Set Low PA was significantly correlated with CBQ Shyness; CTBQ–Set Low Sociability was positively associated with CBQ Low Intensity Pleasure and negatively correlated with CBQ Activity and Impulsivity; and CTBQ–Set Externalizing Behavior was negatively associated with CBQ Shyness and Low Intensity Pleasure. Although not presented here, the correlations between the CTBQ–Set subscales and the CBQ were generally consistent with the pattern for the major scales.

**Discussion**

We developed the CTBQ–Set to provide a measure of preschoolers’ temperament and behavior in a naturalistic setting. In this article, we described the process of scale development and reported data on internal consistency, convergent and discriminant content saturation, interrater reliability, test–retest stability, and construct validity.

The internal consistency of CTBQ–Set major scales and subscales was good to excellent. The high alpha levels for the subscales are especially noteworthy because some subscales include only a few items.
Convergent and discriminant content saturation both ranged from moderate to excellent. All CTBQ–Set subscales correlated more highly with the major scale that they belonged to than with the other major scales by a wide margin. Each major scale’s subscales were moderately to highly intercorrelated, indicating cohesiveness between subscales, and, with one exception, the subscales correlated more highly with the other subscales from the same major scale than with any other subscales. Noncompliance correlated slightly more highly with the anger and emotional regulation deficits subscales of NA than with the other Externalizing Behavior subscales, hyperactivity and aggression. Thus, in our sample, noncompliance appears to have been expressed more frequently by anger and difficulties with regulation of negative affect than through aggression or disinhibition. However, the low frequency of aggressive acts during the observation period may have contributed to the lower correlation between noncompliance and aggression in this sample.

Although there was strong evidence for convergent and discriminant validity, there were a number of significant correlations between subscales from different major scales. These relations reflect the conceptual overlap between several of the constructs of interest (e.g., behavioral impoverishment and low sociability) and are generally consistent with the literature (e.g., Rothbart & Bates, 1998).

The CTBQ–Set major scales demonstrated good interrater reliability for both the first and second home observations. Although not as robust as the major scales, subscale interrater reliability still fell within the adequate to good range.
lems and CBQ Shyness but not CBQ Fear. CTBQ–Set Separation Anxiety/Dependency was significantly associated with CBCL Internalizing Problems (after partialling out the effects of CBCL Externalizing Problems) and with CBQ Shyness. Finally, CTBQ–Set Externalizing Behavior was significantly correlated in the hypothesized direction with CBCL Externalizing Problems and CBQ Activity, Impulsivity, Inhibitory Control, Attentional Focusing, and Anger.

In recent years, several investigators have questioned whether PA and sociability should both be subsumed under the broader construct of extraversion (e.g., Lucas et al., 2000). Our data are consistent with the close relation between these constructs, as CTBQ–Set Low PA and Sociability were the pair of major scales with the highest correlation ($r = .58$). However, there was also evidence for the distinctiveness of these constructs, as each of the Low PA subscales was more highly correlated with each other and with the Low PA major scale than with Low Sociability and its subscales, and the same pattern was evident for the Low Sociability subscales. Interestingly, CTBQ–Set Low PA correlated with CBQ Shyness, and CTBQ–Set Low sociability correlated with CBQ Low Intensity Pleasure. Although this also supports the overlap between the constructs of PA and Sociability, it is possible that trained observers are better able to distinguish between these constructs than are parents.

Consistent with many temperament systems, the CTBQ–Set NA subscales included both internalizing and externalizing components, each of which had somewhat different correlates. The CTBQ–Set sadness/depression subscale correlated more highly with CTBQ–Set Low PA ($r = .53$, $p < .001$), Low Sociability ($r = .39$, $p < .001$), and Separation/Dependency ($r = .34$, $p < .001$) than with Externalizing Behavior ($r = .26$, $p < .05$). In contrast, the CTBQ–Set NA angry/irritable subscale correlated more highly with CTBQ–Set Externalizing Behavior ($r = .58$, $p < .001$) than with CTBQ–Set Low PA ($r = .31$, $p < .01$), Low Sociability ($r = .09$, ns), and Separation Anxiety/Dependency ($r = .10$, ns). Moreover, after partialling out the effects of the other CBCL scale, the CTBQ–Set sadness/depression subscale was correlated with CBCL Internalizing, but not Externalizing, Problems, and the CTBQ–Set angry/irritable subscale was correlated with CBCL Externalizing, but not Internalizing, Problems. Despite these differences, however, the CTBQ–Set sadness/depression and angry/irritable subscales were themselves highly correlated ($r = .75$, $p < .001$). In addition, the CTBQ–Set sadness/depression and angry/irritable subscales were both significantly correlated with CBQ Anger/Frustration ($r = .26$, $p = .01$ and $r = .29$, $p = .005$, respectively), but were only marginally correlated with CBQ Sadness (both $r = .19$, $p < .07$), suggesting that there is some overlap between these two forms of NA. As noted in the literature, it may be difficult to distinguish anger and sadness in preschoolers, as distress responses are not fully differentiated at this stage of development (Mash & Dozois, 1996).

Taken as a whole, these analyses suggest that the CTBQ–Set is a promising measure of preschoolers’ behavior in a naturalistic context. However, this study also has several limitations. First, reported internal consistency and convergent and discriminant content saturation values were used to guide scale development. Thus, it is necessary to cross-validate these results with a different sample. Second, this sample was not large enough to conduct a factor analysis. Although the analyses reported here support the convergent and discriminant validity of the rationally derived CTBQ–Set scales, factor analytic confirmation of the CTBQ–Set structure is needed. Third, we employed a volunteer sample (although approximately half was solicited using commercial mailing lists) that consisted predominantly of working lower middle class, Caucasian, two-parent families. Although these demographics are characteristic of the area in which the study was conducted, and both internalizing and externalizing temperamental characteristics have shown stability across ethnically and socioeconomically diverse populations (Achenbach, 1995; Huebeck, 2000), it is important to evaluate the CTBQ–Set in more representative and diverse samples. Fourth, studies using clinical samples are necessary to evaluate the CTBQ–Set’s clinical utility. Fifth, it is important to determine if the CTBQ–Set functions equally well in other settings, such as at school. Sixth, it is also important to examine the stability of the CTBQ–Set over longer time intervals. Finally, the upper age limit of the CTBQ–Set is unclear. Although the major constructs tapped by the CTBQ–Set are relevant for older children (Shiner, 1998), the specific behaviors used to measure those constructs may change over the course of development.

The CTBQ–Set may be useful to other researchers in several ways: (a) it can be used to augment parent-report measures or more structured laboratory paradigm derived data; (b) the broad set of CTBQ–Set items allow the possibility of constructing scales for other potentially important constructs; (c) it can be employed as an ipsative measure, yielding an individual profile of salient behaviors for each participant; and (d) the CTBQ–Set can be used to conduct cluster analyses of children based on these individual profiles. The CTBQ–Set was not designed for use in clinical practice, as it is probably too time consuming to be cost-efficient to practitioners. However, the CTBQ–Set yielded reliable results with trained graduate and undergraduate student observers, suggesting that a trained nonprofessional staff member could use it to collect data on temperament and behavior for the clinician’s use.
References


**APPENDIX**

**CTBQ–Set Scales and Subscales**

**Low PA**

**Affective Constriction**

1R. Predominant mood is happy (vs. apathetic, serious, depressed, angry, or anxious)

2. Emotionally constricted; child’s affect is characteristically indifferent, disengaged or flat—note facial movements, gestures, vocalizations (vs. child is affectively expressive)

3. Child is serious and solemn most of the time (vs. light-hearted, fun-loving and playful, gets a “kick” out of life)

4R. When child exhibits positive affect, it is very strong—exuberance, “lights up,” shrieks or doubles over with laughter (vs. little or no positive affect)

5R. Laughs easily with caregiver, observer, or familiar adults (vs. rarely laughs with anyone)

6. Smile seems forced, limited to mouth, or mouth entirely closed; fades quickly or rarely smiles at all (vs. smile is broad, involving entire face and eyes)

7. Child has a high threshold for positive affect; it takes a lot of work to elicit even moderate positive affect (vs. threshold for positive affect is low; doesn’t take much to elicit smiling, laughing)

**Behavioral Impoverishment**

8. Is apathetic, uninterested, unresponsive to activities that most children would find interesting or pleasurable (vs. interested, responsive)

9R. Spontaneously expresses enjoyment of accomplishing or achieving (vs. does not express enjoyment of accomplishing or achieving even when prompted)

10. Appears lethargic, listless, lacking in energy, easily tired, low stamina, lies around house (vs. is lively and energetic)

11. Speech has monotonous or flat tone (vs. speech is expressive or animated)

12. Unoccupied/unengaged—requires adult encouragement to keep occupied or just sits around and wanders aimlessly (vs. rarely unengaged—initiates activities or keeps occupied on own)

13. Play is repetitive, concrete, lacks imagination—does not combine several objects in play, plays with one at a time when several are available and could be used together (vs. creative/imaginative in play)

14. Child appears dreamy or “spaced out” (vs. child is generally alert and interested in what is going on in immediate environment)

15. Much of time is spent engaged in passive activities (i.e., watching TV)—rather than activities requiring greater cognitive or physical effort (vs. drawn to challenging tasks and activities requiring at least moderate physical or cognitive effort)

16R. Curious and exploring, wants to explore new toys and activities, asks caregiver or observer about events or objects in the environment

17. Lacks self-confidence; puts self down or criticizes self—i.e., “I’m no good at [activity]”; or verbalizes feelings of helplessness or lack of self-efficacy—i.e., “I can’t do it” (vs. appears sure of him/herself)

**Low Appetitive Behavior**

18R. Physical movement is rapid, vigorous, forceful (vs. movement is slow, sluggish)

19R. Is eager to demonstrate songs, games, abilities, or other behavior—asks caretaker to watch him/her perform (vs. does not demonstrate abilities at all)

20. Child loses interest and disengages when an activity looks like it might be difficult—low persistence (vs. child persists despite difficulties; does not give up easily; may calmly request help)

21R. When child is anticipating participating in an exciting activity or playing with a new toy, he/she becomes visibly excited—trembles or jumps up and down, can’t stop asking when it will be time or talking about it

22R. Has fun, even during routine activities

23R. Child responds with enthusiasm to invitations to play or engage in a pleasurable activity

24R. When child sees something he/she wants to do, he/she goes after it right away

**NA**

**Sadness/Depression**

25. Child exhibits droopy or sad posture—muscle tone is flaccid, slumped shoulders
26. Predominant affect/mood is sad or depressed; child exhibits mild but pervasive sadness for much of the observation; or frequent but brief sadness (vs. child does not, or if so, sadness is momentary, and clearly in response to a specific event)

27R. Child does not cry or appear tearful (vs. cries or appears to be on the verge of tears at least several times)

28. Tends to pout or whine (vs. no evidence of pouting or whining)

Angry/Irritable

29. Becomes angry with caregiver—note facial expression, directed anger, tantrumming (vs. never even mildly angry at caregiver)

30R. Child is not upset by minor frustrations; is rarely even mildly irritated or angry (vs. overreacts to minor frustrations; is easily irritated or angered)

31. Appears sullen or resentful (vs. no evidence of sullenness or resentment)

Anxiety/Fear

32. Nervous or anxious; note subtle postural cues, such as body tension or vigilance, as well as verbal statements; worries, startles easily, or has nervous habits or tics, such as biting nails, twisting hair, pulling on clothes)

Oversensitivity

33. Unperturbed by teasing, put-downs, or criticism (vs. feelings are easily hurt)

Emotional Regulation Deficits

34R. Child recovers quickly from distress, e.g., anger, sadness, fear (vs. not easily distracted from distress or takes a long time to recover from distress)

35. Has rapid shifts in emotional state or mood; is emotionally labile (vs. emotions don’t change often or easily)

36R. Is calm, relaxed, easygoing (vs. is high strung, tense, prone to having “meltdowns”)

Low Sociability

Low Sociability

37R. Initiates interaction with observer (vs. does not initiate interactions)

38R. Initiates interaction with siblings or other familiar children (vs. does not initiate interactions)

39R. Acts to maintain social interaction, includes observer after becoming familiar; child engages others in conversation, asks and answers questions, makes directed comments

40. Is shy and reserved, even with familiar individuals (vs. outgoing, gregarious)

41. Is indirect and hesitant in addressing observer, even after being warmed up (vs. addresses observer directly, after being warmed up)

44R. Engages people to the exclusion of object play and exploration, seems to lose interest in activities once left alone to play (vs. child spends much of the observation period engaged in solitary play or activities; appears to prefer playing alone)

45. Child seems uninterested in, and largely ignores, adults who visit the home—even after a warm-up period

Fear of Novel Situations

42R. Child quickly gets used to, and warms up to, people who initially elicited shyness

43. Child retreats to caregiver, withdraws, or hides when new people visit the home; may avert gaze when meeting or interacting with an unfamiliar person (vs. greets unfamiliar people, including observer, spontaneously)

Separation Anxiety/Dependency

46R. Does not attend to or monitor changes in caregiver’s location or activities (vs. is aware—hypervigilant—about changes in caregiver’s location or activities)

47. Bouts of exploration or play away from caregiver are brief or attempts to follow when caregiver moves away

48. Becomes distressed by separation from caregiver—i.e., caregiver leaving room; may cling to caregiver

49. Is bolder or more confident in play when caretaker is nearby (vs. no change when caregiver is absent)

50. Requests caregiver’s help in completing tasks child should be able to do on his/her own

51. Child becomes disorganized—play is less goal oriented—in caretaker’s absence; seems to require contact with caregiver to maintain interest or for emotional regulation (vs. quality of play does not change when caregiver is absent)
52R. Child seeks to be independent and autonomous (vs. child is clingy, dependent, helpless)

**Externalizing Behavior**

**Hyperactivity–Impulsivity**

53. A risk taker; engages in daring or potentially hazardous physical activity—climbing and jumping from high places (vs. physically cautious; child will not try activities that are even mildly risky or physically dangerous)
54. Average level of physical activity is high; child is always on the go or moving around; may run or skip instead of walk (vs. low level of physical activity)
55R. Able to sit still for extended periods of time, during quiet activities such as TV or story time (vs. restless, fidgety, difficulty sitting still during quiet activities)
56R. Generally quiet, speaks softly (vs. loud, shouts, makes a lot of noise)
57R. Explores objects thoroughly (vs. contact with objects is brief/superficial)
58R. Child becomes engrossed in his/her play or persists in one activity for a long time without encouragement (vs. doesn’t sustain interest)
59. Leaves many activities or tasks uncompleted in order to do something that appears more novel or desirable
60R. Not impulsive; thinks about actions/behaviors before doing or saying something (vs. impulsive; does not seem to think about his/her actions/behavior before doing or saying something)
61. Easily distracted by what is going on around him/her; has difficulty sustaining attention (vs. is attentive and able to concentrate)
62R. Alert and responsive when spoken to directly (vs. often does not seem to listen when spoken to)
63. High levels of aimless physical activity—non-directed, non-essential movement—or vocal behavior that is frequently non-directed and non-social (vs. physical activity and speech are generally controlled and goal-directed)

**Noncompliance**

64R. Compliant, obedient; child follows caregiver’s suggestions readily even when they are suggestions rather than commands (vs. defies or refuses to comply with adults’ requests or rules)
65. Characteristically pushes and tries to stretch limits; tries to see what he/she can get away with
66. Argues with caregiver when given instruction

67R. When caregiver says “no” or punishes child, he/she stops misbehaving—at least at the time (vs. child continues misbehaving)
68. Impatient; talks out of turn, often interrupts others, has difficulty waiting to take turns in games or goes out of turn
69. Child wants to be the center of adults’ attention; if caregiver or observer is busy or talking to someone, child interrupts or engages in negative attention-getting behavior
70. Jealous; does not accept caregiver’s or observer’s attention to others

**Aggression**

71R. Not physically aggressive (vs. physically aggressive; hitting, biting, kicking, scratching, hairpulling, poking, throwing objects at people)
72. Is verbally aggressive, such as name calling, insulting, hostile teasing, or other aggressive verbal behavior
73. Does not adapt physical play to avoid hurting others—may not appear to be directly aggressive, but child’s actions nonetheless result in injury, pain, or discomfort to the person who is the recipient; physically intrudes on others’ space in an aggressive manner—climbs on caregiver and in the process kicks, pulls on hair, pokes or scratches him/her
74. Aggressive with toys or other objects (hits, bangs, kicks, or throws them)
75R. Generally careful and gentle in playing with, or handling, toys or other objects

**Miscellaneous Problems**

*76. Is very upset by negative evaluations or disapproval from caregiver; child may cry or become distressed when corrected by parent, even if correction was not made in a harsh or particularly negative tone (vs. does not appear affected by caregiver’s corrections or disapproval)

77. Child rejects caregiver’s attempts to soothe him/her (vs. readily accepts soothing)
78. Solicits and enjoys playful interaction with caregiver
79. Child appears sad or dejected when he/she is not the center of attention (vs. mood does not change when not the center of attention)
80. Is able to delay gratification (vs. unable to delay gratification; demanding when required to

*Miscellaneous items removed from a priori scales for psychometric reasons (76 and 77, NA; 78, Low Sociability; 79, Separation Anxiety; 80, Hyperactivity)
wait for satisfaction; may fuss and persist until caregiver does what the child wants right away; makes requests in loud, insistent tone of voice)

81. Empathic; is able to see how others feel; can put him/herself in their place

82. Has a hard time making up his/her mind; changes his/her mind a lot (vs. decisive; sure of what he/she wants)

83. Is bossy; tries to dominate other people (vs. submissive)

84. Child’s observations and requests are easy to understand; is verbally fluent (vs. speech is unclear, difficult to understand)

85. Is physically attractive (vs. physically unattractive)

86. Is agile and well-coordinated; gross motor control is smooth and coordinated (vs. clumsy; physically awkward; poor coordination)

87. Fine motor manipulation is not skillful—draws very poorly; has difficulty placing blocks or puzzle pieces in proper position

88. Is visibly deviant from peers in appearance, size, or physical condition (vs. normal in appearance, size, and physical condition)

89. Child complies readily when observer indicates—verbally or by disengaging—that he/she cannot play with child (vs. child continues to try to play with observer despite observer indicating that he/she cannot play with the child at that moment)

90. Child appears coy, flirtatious, seductive—not necessarily in a sexualized manner— or manipulative with caretaker, observer, or other adults

Received April 4, 2001
Accepted May 20, 2002