Parents of small children commonly experience child management difficulties, and health care professionals are in a unique position to address parental concerns (Howard, 1991). Behavior problems are especially prevalent in toddlers as parents begin to establish routines and expect compliance (Mash & Johnston, 1983). For example, Stallard (1993) found that 66% of mothers surveyed expressed two or more concerns about the behavior of their 3-year-old child. Common problems encountered by parents of toddlers include aggressiveness, noncompliance, temper tantrums, shyness, and problems related to sleeping, shopping, and mealtime routines. Although it is typically assumed that behavior problems in toddlers are transient and will remit before kindergarten (Miller & Scarr, 1989), as many as 50% will continue to exhibit significant behavior problems into elementary school (Campbell, Ewing, Breaux, & Szumonski, 1986; Richman, Stevenson, & Graham, 1982). Early detection and treatment is by far the most efficient and effective method of service delivery (Bennett & Guralnick, 1991).

The importance of behavioral screening of preschoolers as a routine part of pediatric primary care has been emphasized repeatedly (Christophersen, 1984; Lavigne et al., 1993; Stancin & Palermo,
of behavioral concerns has been the lack of valid family competency (Kemper, 1992; Shorr, 1995). It is evident, however, that focus on parental depression and distress of child behavior (Kemper, 1992). It is evident, however, that focus on parental depression and distress are important to recognize and address. The prevalence of many of these risk factors is highest in mothers of preschool children; thus, these families represent a particularly high-risk group. Unfortunately, the assessment of psychosocial risk factors is even less frequently conducted by health care professionals than assessment of child behavior (Kemper, 1992). It is evident, however, that focus on parental depression and distress can mitigate parenting difficulties and enhance family competency (Kemper, 1992; Shorr, 1995).

One obstacle to effective and efficient detection of behavioral concerns has been the lack of valid and reliable screening measures for pediatricians to employ (Stancin & Palermo, 1997). Existing instruments are either too lengthy to serve as a screening measure or too difficult to score and interpret for efficient use in well-child care. For example, the Child Behavior Checklist (CBCL) is a well-established instrument considered too lengthy and time-consuming to score in a pediatric setting (Jellinek & Murphy, 1990; Stancin & Palermo, 1997). Furthermore, currently available questionnaires are generally intended for elementary school children and lack items relevant to toddlers (Mouton-Simien, McCain, & Kelley, 1997). Stancin and Palermo (1997), in their critique of behavioral screening practices in pediatric settings, state that “preschool age children are under represented in studies of behavioral screening” (p. 185).

A valid screening measure is effective in making accurate decisions to refer children and parents who need services and minimizes the chance of falsely identifying children (Ireton, 1990). Furthermore, screening questionnaires must be psychometrically sound yet practical, easy to complete, and quick to score (Jellinek & Murphy, 1990). A brief screening questionnaire is often used by pediatricians to obtain children’s health history. Employing a similar screening questionnaire may prove to be a valid and efficient first step of early detection of behavior problems and other risk variables (Christophersen, 1986, Lavigne et al., 1993; Stancin & Palermo, 1997).

The Toddler Behavior Screening Inventory (TBSI) was developed to provide a developmentally appropriate screening questionnaire for health care professionals to use as a part of well-child care (Mouton-Simien et al., 1997). The instrument contains 40 items judged to be frequent sources of concern for mothers of toddlers. The instrument was developed through a series of studies. Study 1 generated a large pool of times (n = 93) from mothers of preschoolers. Study 2 reduced the list of items to a subset of items judged to be frequent and problematic by mothers (n = 312). Items were retained if 30% of the sample indicated the behavior frequently occurred and 10% of the sample indicated that the behavior was problematic for their child. Study 3 examined the psychometric properties of the scale. Initial support for the reliability, validity, and utility of the scale was found (Mouton-Simien et al., 1997). Furthermore, the questionnaire was highly correlated with a more lengthy, alternate instrument, the CBCL.
The primary purpose of this study was to obtain additional data on the psychometric properties of the TBSI and extend its use to a clinical sample. Additionally, we evaluated whether differences across age groups as well as other demographic variables existed. Another primary purpose was to evaluate the utility of the TBSI in differentiating clinical from nonclinical samples. A secondary purpose was to examine the contribution of assessing maternal risk factors to identifying the clinical sample. This purpose was intended to evaluate whether data on factors such as depression and maternal distress contributed significantly to the prediction of clinical status above and beyond that obtained with TBSI scores only.

Method

Participants

The sample consisted of 362 mothers with toddlers between the ages of 12 and 41 months. Both a nonclinical sample (N = 312) and a small clinical sample (N = 50) were recruited from a variety of settings including pediatricians’ offices, child care centers, and a local charity hospital. Although similar to the sample used in Mouton-Simien et al. (1997), the sample was somewhat more heterogeneous and was comprised of an entirely different set of participants. Nonclinical participants were grouped according to age: 12–23 months (37%), 24–35 months (41%), and 36–42 months (22%). Nonclinical participants were 51% female and 49% male. Mothers were Caucasian or Asian (83%) or African American (17%) married (78%) women. Mothers came from varying income levels with 23% having a household income of less than $25,000, 27% having incomes between $25,000 and $50,000, and 50% having household earnings over $50,000 yearly. With regard to maternal education, 20% had a high school diploma or less, 32% had some college, and 32% completed college.

The clinical group consisted of mothers of toddlers referred by pediatricians for outpatient psychological treatment. The pediatricians who referred the clinical group were on staff at the local charity hospital or had private practices from which the nonclinical sample was drawn. Toddlers in the clinical group were 46% female and 54% male. With regard to the toddlers’ age, 12% were between 12 and 23 months, 54% were between 24 and 35 months, and 34% were between 36 and 42 months. Mothers of toddlers in the clinical group were very similar in most respects to mothers in the nonclinical group. They were primarily Caucasian or Asian (72%) married (64%) women. The remainder were African Americans. Over half of the mothers had completed high school and attended some college (58%). Household incomes for the clinical sample were lower than for the large, nonclinical sample. Household income was below $25,000 for 48% of the clinical sample as compared to 23% of the large, nonmatched, nonclinical sample.

All toddlers were screened using the Denver II (Frankenburg & Dodds, 1990), a developmental screening instrument, and were found to be developing within normal limits. Signed written consent for voluntary participation in the study was obtained from each mother.

Measures

Demographic Questionnaire. A brief questionnaire was completed to obtain information on socioeconomic status (SES), marital status, household composition, and other demographic characteristics. Ratings of SES were computed for each participant based on Hollingshead’s two-factor theory, which utilizes parental education and occupational status (Hollingshead & Redlich, 1958).

A question regarding past and present psychological services was included as one method of ensuring integrity of group status. Families, parents, or children currently receiving psychological services were not included in the nonclinical group.

Toddler Behavior Screening Inventory. The TBSI is a 40-item checklist that assesses the frequency and intensity of toddler behavior problems (Mouton-Simien, 1990; Mouton-Simien et al., 1997). Sample items include refusal to share, temper tantrums, irritability, crying at bedtime or during the night, throwing objects, refusal to try new food, and clinging to parents. Mothers are instructed to respond to each item according to their child’s behavior within the previous month. For each item, respondents rate the frequency on a three point scale (O = Not True, 1 = Somewhat True/Sometimes True, and 2 = Very True/Often True). Each item is also rated on whether or not the behavior is problematic by a participant circling “yes” or “no.” The scale yields two scores, a total frequency and a total problem score.
Frequency Scale scores potentially could range from 0 to 80 and Problem Scale scores from 0 to 40. The two dimensions were evaluated because items may occur frequently and yet not be considered a problem by a mother given her expectations for her child and other developmental considerations.

*Achenbach Child Behavior Checklist (CBCL/2–3)* (*Achenbach, 1992*). The 99-item CBCL/2–3 is a parent-report measure designed to assess behavioral and emotional problems in children ages 2 to 3. The scale has an extensive literature supporting its psychometric integrity (Achenbach, 1992). The CBCL/2–3 contains six first-order factors and two second-order factors (Achenbach, 1992). For the purpose of this study, the Internalizing, Externalizing, and Total Problem factor scores were used.

The following four questionnaires were used to measure various parent “risk” variables.

**Beck Depression Inventory (BDI).** The BDI is a well validated, widely used 21-item self-report measure of depression in adults (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961). The Beck was used as a means of measuring depressive symptoms in mothers.

**Dyadic Adjustment Scale (DAS).** The DAS is a 32-item, widely used questionnaire that evaluates general marital satisfaction. Adequate reliability and validity have been demonstrated (Spanier, 1976). The total score of the DAS was entered into the analyses.

**Mothers Activity Checklist–Revised (MAC).** To obtain a measure of maternal isolation and intensity of negative daily events, the 45-item MAC was given. The MAC was developed to assess a range of maternal setting events both pleasant and unpleasant (Kelley & Carper, 1988). One of the four factors, the Unpleasant Events factor, was used to measure life stressors. This factor was chosen as it has been found to be associated consistently with negative maternal perceptions of children.

**Interpersonal Support Evaluation List (ISEL).** The ISEL is a 40-item measure of self-reported social support availability (Cohen & Willis, 1985). Adequate reliability has been reported (Cohen & Willis, 1985). The ISEL contains four factors, but only the Tangible Support scale score was used in the analyses.

**Procedure**

Participating mothers were told that the study was investigating common problems experienced during toddlerhood. After giving voluntary written consent, the participants completed the following questionnaires: a demographic questionnaire, the TBSI, CBCL, BDI, DAS, ISEL, and MAC. Assistants encouraged participants to ask questions regarding the questionnaire completion. After questionnaire completion, the Denver II was administered (Frankenburg, & Dodds, 1990).

**Results**

We conducted data analyses according to the following plan. First, the reliability and concurrent validity of the TBSI were assessed. Second, means and standard deviations were calculated. In addition, demographic and developmental differences between participants were assessed. Third, the effect of clinical status on TBSI scores was evaluated. Finally, we analyzed the ability of maternal risk factors to discriminate clinical from nonclinical samples beyond that obtained with TBSI scores.

**Reliability**

The data revealed that both the TBSI frequency and problem scale were internally consistent. Cronbach’s alpha equaled .84 for the frequency scale and .85 for the problem scale. Item-total correlations ranged from .18 to .59 for the frequency scale and .27 to .63 for the problem scale. Adequate test-retest reliability was obtained when 31 mothers were re-administered the TBSI after a 2-week interval ($r = .83$ for both frequency and problem scale).

We evaluated whether each item continued to meet the criteria for inclusion utilized by Mouton-Simien et al. (1997). In Mouton-Simien et al., item retention criteria were (a) a frequency scale mean of .60 or greater (i.e., 10% or greater of the sample indicated frequent occurrence of the behavior), (b) a problem scale mean of 10 or greater (i.e., 10% or greater of the sample indicated they considered the behavior was problematic, and (c) item-total correlation coefficient of .20 or greater for the frequency and problem scales. Using the same criteria as Mouton-Simien et al., we found that all items except “speaks poorly” met criteria. The “speaks poorly” item also did not meet exclusion criteria in Mouton-Simien et al., but was retained due to its clinical relevance to toddlers. Thus, the TBSI contains items that occur fairly frequently and are rated as somewhat problematic by mothers of nonclinical toddlers.
Developmental Differences

Univariate ANOVA yielded significant differences between toddler age groups on the frequency scale, $F(2, 321) = 6.42, p < .002$, but not the problem scale. Specifically, mothers of 1- and 2-year-old toddlers rated their toddlers as exhibiting higher rates of the behaviors described in the scale than mothers of 3-year-olds. However, mothers did not perceive the behaviors as significantly more problematic.

Specific Problematic Behaviors

In evaluating responses to individual items, the behaviors that most mothers perceived as problematic included whining (35%), tantrums (36%), throwing objects (28%), and wandering away from parent (31%). Although many of the behaviors were rated as frequent at all age groups, some were rated higher at different developmental levels. Mothers of younger toddlers rated behaviors associated with the child’s physical safety and behavioral disinhibition as occurring at the highest rates and being the most problematic. For example, the most frequently rated behaviors by mothers of 1-year-olds were putting inappropriate things in mouth (18%), climbing on furniture (23%), and pulling things out of cabinets (19%). Mothers of older toddlers reported more frequent occurrence of noncompliant behaviors (e.g., disobeying instructions [57%], refusal to nap [23%]) and behaviors involving language and interactions with others (e.g., talking back [26%], refusal to share [20%]).

Effects of Clinical Status on TBSI Scores

We used univariate ANOVAs to examine differences in TBSI scores between clinical and nonclinical participants. A matched sample of 50 nonclinical participants was chosen from the total sample ($N = 325$) and used in the analyses. Specifically, 50 nonclinical participants were matched on child age, gender, and race and maternal SES. These characteristics were matched to ensure that the two groups were similar. As seen in Table I, significant differences between clinical and nonclinical groups were obtained across all ages with the exception of the 1-year-old toddlers’ scores on the problem scale. That is, 2- and 3-year-old clinical participants scored sig-
significantly higher than nonclinical participants on both scales of the TBSI. With regard to gender differences, both males and females in the clinical group scored significantly higher than males and females in the matched nonclinical group, as seen in Table I.

Discriminant analyses were conducted on the matched sample to provide additional information on how well the TBSI Problem and Frequency Scale scores classified participants into clinical and nonclinical groups. The frequency scale correctly classified 76% of the cases into either the clinical or nonclinical groups. The problem scale resulted in 73% of the 100 participants being correctly classified. When both scales were used, 82% of the participants were classified correctly. Thus, scores on the scale were quite good at differentiating families who were referred for behavioral counseling by their pediatrician from those who were not referred.

**Suggested Cut-off Scores**

The results obtained here with the nonclinical sample and those obtained in the previous study (Mouton-Simien et al., 1997) were quite equivalent. For the nonclinical sample in the current study, the overall mean for the Frequency scale was 24 (SD = 9) and 26 (SD = 9) in the previous study. The same mean for the clinical sample presented here was 39 (SD = 9). The overall mean for the nonclinical sample on the Problem Scale was 6 (SD = 5) in our study and 7 (SD = 6) for the previous study. The Problem Scale mean for the clinical sample was 17 (SD = 7). Based on these data, we suggest that cut-off scores of 5 on the Frequency Scale and 10 on the Problem Scale be utilized at this time and until further data are obtained. These numbers are suggested as they represent scores about one standard deviation above the nonclinical sample mean. Additionally, 45 of the 50 clinical participants had either a Problem Scale or Frequency Scale score above the cut-off and therefore would have been identified as in need of further assessment. Using only the Frequency Scale score, we would have identified 37 participants in the clinical sample. Finally, using the Problem Scale score, we would have identified 42 participants in the clinical sample.

We caution against the overreliance on these suggested cut-off scores given the very preliminary nature of the data. We recommend that any items endorsed as “problematic” receive further inquiry.

**Relationship with Risk Factors**

We conducted discriminant analyses to determine if scores on the measures of risk significantly enhanced the TBSI’s ability to classify clinical from nonclinical participants. The TBSI Scale scores and maternal risk factors were used as predictor variables. The procedure revealed that the frequency scale, problem scale, BDI, Frequency of Unpleasant events and Tangible Support factor of the ISEL combined to accurately classify 88% of the participants (Wilks’s = .67, p < .0001). Thus, the addition of the measures of risk improved the classification of participants from 82% to 88% correctly classified, using a combination of child and maternal variables. This suggests that clinical status is slightly better classified when maternal risk factors are included in the analyses. In particular, maternal depression, negative life events, and a lack of social

---

**Table I.** Scores for Clinical and Matched Normal Groups on the Frequency and Problem Scales

<table>
<thead>
<tr>
<th></th>
<th>Clinical group (n = 50)</th>
<th>Matched group (n = 50)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Freq. scale M (SD)</td>
<td>Prob. scale M (SD)</td>
</tr>
<tr>
<td>All participants</td>
<td>39.04 (8.80)</td>
<td>17.22 (7.40)</td>
</tr>
<tr>
<td>Toddler age (months)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12–23</td>
<td>39.00 (8.17)</td>
<td>8.83 (3.76)</td>
</tr>
<tr>
<td>24–35</td>
<td>40.00 (8.78)</td>
<td>17.56 (6.74)</td>
</tr>
<tr>
<td>36–42</td>
<td>37.79 (9.31)</td>
<td>19.47 (7.46)</td>
</tr>
<tr>
<td>Toddler gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>38.38 (8.46)</td>
<td>15.19 (6.08)</td>
</tr>
<tr>
<td>Females</td>
<td>39.75 (9.29)</td>
<td>19.42 (8.17)</td>
</tr>
</tbody>
</table>

* p < .05.  
** p < .0001.
support in helping with routine tasks were associated with the increased likelihood that mothers and their toddlers would be referred for behavioral counseling above and beyond that indicated by the TBSI scores alone.

Discussion

The TBSI is intended as a brief screening measure for assessing commonly occurring toddler behavior problems. The TBSI can be completed in 10 minutes, is simple to score, and is economical for pediatricians or child care providers to give to large numbers of children. The results of the current study support the reliability and validity of the instrument. Consistent with Mouton-Simien et al.’s (1997) results, our results indicated that the TBSI is reliable and internally consistent. Further, this study found that less educated, African American, more impoverished mothers found their toddler’s behavior to be more distressing than did their counterparts. Finally, type and severity of problem behaviors endorsed varied with the age of the child.

A primary purpose of this study was to evaluate the utility of the TBSI in differentiating clinical and nonclinical samples. The results indicated that mothers who were referred by their pediatricians for behavioral counseling reported significantly more problematic behaviors than mothers of nonreferred toddlers. Furthermore, the discriminant function analysis revealed that 82% of the participants were correctly classified as either clinical or nonclinical participants using the two scales of the TBSI. These results suggest that the TBSI is a good discriminator of clinical status. Another goal of the present study was to evaluate whether assessment of maternal risk factors further contributed to the classification of referred and nonreferred participants. Indeed, maternal depression, inadequate tangible social support, and negative life events contributed to the classification of participants. Specifically, inclusion of maternal risk factors increased the correct classification of participants from 82% to 88%. However, further research is needed on the sensitivity and specificity of these variables in aiding in the classification of participants. Given the relatively small contribution of the maternal measures, above and beyond that obtained with the TBSI, it may be unnecessary or inefficient to include these measures for screening. However, additional research regarding the importance of screening maternal factors is needed.

Although this study provides preliminary evidence regarding the TBSI’s utility in discriminating clinical and nonclinical groups, there are several limitations. First, this sample was somewhat homogeneous in race, educational level, and marital status. Evaluation of the TBSI using a more heterogeneous sample, including more highly educated, married, African American participants, is needed. Further evaluation of the relationship between maternal risk factors and child clinical status is needed. Specifically, mothers in this sample were not considered clinically distressed and a clinical sample of mothers was not employed. Follow-up studies using the TBSI to examine the relationship between development status and problematic behavior are needed. Finally, the validity of implementing the TBSI into a pediatric practice and the accuracy of identification of significant problems by physicians should be studied.

Although pediatricians have a unique opportunity to detect, guide, and refer preschoolers with behavioral problems, research has shown that many children exhibiting significant behavior problems go undetected (Costello et al., 1988; Lavigne et al., 1993). Reasons for overlooking preschoolers’ behavior problems include pediatricians’ lack of training in behavioral issues, difficulty of assessing due to limited time with patients, and lack of available screening instruments for young children. However, the importance of early intervention is critical to preventing long-term behavior problems at home, school, and with peers (Bennett & Guralnick, 1991; Campbell et al., 1986; Prior, Smart, Sanson, Pedlow, & Oberklaid, 1992). The TBSI appears to be a promising instrument for screening common behavior problems in toddlers. It can be completed and scored quickly and thus shows potential for use in pediatric practice. At this time, we recommend that cut-off scores of 35 for the Frequency Scale and 10 for the Problem Scale be used based on the findings obtained here and those of the previous study. However, we caution against overreliance on cut-off scores and emphasize that further inquiry, advice, and possible referral for treatment be considered for any parent who endorses an item as problematic.

Received November 24, 1997; revisions received March 30, 1998; accepted January 7, 1999
References


