Published in final edited form as:

J Abnorm Child Psychol. 2013 August ; 41(6): 851–863. doi:10.1007/s10802-013-9738-8.

Identification of Developmentally Appropriate Screening Items for Disruptive Behavior Problems in Preschoolers

Christina R. Studts, Ph.D. and

Department of Behavioral Science, University of Kentucky, 101 Medical Behavioral Science Building, Lexington, KY 40536-0086, Office: 859-323-1788, Fax: 859-323-5350

Michiel A. van Zyl, Ph.D.

Kent School of Social Work, University of Louisville, Oppenheimer Hall, Louisville, KY 40292

Christina R. Studts: tina.studts@uky.edu

Abstract

OBJECTIVE—Screening preschool-aged children for disruptive behavior disorders is a key step in early intervention. The study goal was to identify screening items with excellent measurement properties at sub-clinical to clinical levels of disruptive behavior problems within the developmental context of preschool-aged children.

METHOD—Parents/caregivers of preschool-aged children ($N = 900$) were recruited from four pediatric primary care settings. Participants (mean age = 31, $SD = 8$) were predominantly female (87%), either white (55%) or African-American (42%), and biological parents (88%) of the target children. In this cross-sectional survey, participants completed a sociodemographic questionnaire and two parent-report behavioral rating scales: the PSC-17 and the BPI. Item response theory analyses provided item parameter estimates and information functions for 18 externalizing subscale items, revealing their quality of measurement along the continuum of disruptive behaviors in preschool-aged children.

RESULTS—Of 18 investigated items, 5 items measured only low levels of disruptive behaviors among preschool-aged children. The remaining 13 items measured sub-clinical to clinical levels of disruptive behavior problems (i.e., $> 1.5 SD$); however, 5 of these items offered less information, suggesting unreliable measurement. The remaining 8 items had high discrimination and difficulty parameters, offering considerable measurement information at sub-clinical to clinical levels of disruptive behavior problems.

CONCLUSIONS—Behaviors measured by the 8 selected parent-report items were consistent with those identified in recent efforts to distinguish developmentally typical misbehaviors from clinically concerning behaviors among preschool-aged children. These items may have clinical utility in screening young children for disruptive behavior disorders.

Keywords

disruptive behavior disorders; preschool; primary care; developmental; screening

Disruptive behavior problems including aggression, rule-breaking, defiance, and cruelty manifest not only in adolescents and older children, but also in very young children. Preschool-aged children who are “early starters” with respect to such behaviors are at high

Correspondence to: Christina R. Studts, tina.studts@uky.edu.

The authors declare that they have no conflicts of interest.

risk of a continuing developmental pathway of antisocial behaviors. Multiple longitudinal studies have demonstrated the stability and entrenchment of disruptive behaviors first assessed at preschoolage (Harvey, Youngwirth, Thakar, & Errazuriz, 2009; Keenan et al., 2011; Pierce, Ewing, & Campbell, 1999). Recent prevalence estimates of disruptive behavior disorders among children ages 3 to 5 in the United States (U.S.) suggest that 8% to 13% meet Diagnostic and Statistical Manual (DSM-IV; American Psychiatric Association, 1996) diagnostic criteria for Oppositional Defiant Disorder (ODD; Lavigne, Lebailly, Hopkins, Gouze, & Binns, 2009), while 3% to 6% meet the diagnostic criteria for Conduct Disorder (CD; Egger & Angold, 2006; Keenan, Shaw, Walsh, Delliquadri, & Giovannelli, 1997; Kim-Cohen et al., 2005). However, it is likely that more than one in five children exhibit sub-threshold symptoms (Costello & Shugart, 1992), increasing risk for development of later problems. The overwhelming majority of children and families experiencing these problems do not receive specialized services (Fantom, MacDonald, & Harvey, 2008; Kataoka, Zhang, & Wells, 2002; Lavigne et al., 2009), highlighting the importance of early identification and intervention efforts.

Pediatric primary care is an ideal setting for screening and early identification efforts (Agency for Healthcare Research and Quality, 2002), especially for children not yet attending school. While the significance of behavioral issues in primary care settings has been recognized, primary care physicians have struggled with persistent under-identification of children in need of services (Costello & Edelbrock, 1985; Fantom et al., 2008; Lavigne et al., 1993). To improve rates of identification in pediatric primary care, standardized screening approaches using reliable and valid instruments may be helpful (Halfon, Regalado, McLearn, Kuo, & Wright, 2003; Hill, Lochman, Coie, & Greenberg, 2004), and development of very brief, highly informative parent-report screening tools for preschool-aged children is crucial.

Most instruments measuring disruptive behaviors in children are inappropriate for screening purposes in primary care settings, due to (a) excessive time required for administration, scoring, and interpretation; (b) prohibitive costs; and (c) development with non-representative norming samples. In contrast, brief, easily scored, freely available instruments such as the Pediatric Symptom Checklist-17 (PSC-17; Gardner et al., 1999) and the Behavior Problems Index (BPI; Peterson & Zill, 1986; Zill, 1990) may be valuable tools for pediatric primary care. Each of these instruments includes subscales intended to measure disruptive behavior problems and is intended for use with general populations of children. Because they include items tapping associated conditions and risk factors for disruptive behavior, rating scales such as the PSC-17 and BPI may increase screening sensitivity as compared to symptom checklists comprising diagnostic criteria only, according to the principles of public health screening (Wilson & Jungner, 1968). However, the measurement properties of the PSC-17 and the BPI with regard to early identification efforts with very young children are unknown, despite the importance of targeting children in the preschool age range for early identification purposes. In addition, even these brief instruments may be too long for use in primary care settings. Recent advances have been made in developing “ultra-brief” (i.e., 5 items) screening tools for several psychosocial problems in adults in primary care, including anxiety (Berle et al., 2011), depression (Arroll, Goodyear-Smith, Kerse, Fishman, & Gunn, 2005), and drug and alcohol misuse (Bradley et al., 2007). Indeed, 4–6 items may be sufficient for measuring most constructs (Hinkin, 1998; Tabachnick & Fidell, 2001).

Further complicating early identification through screening is the developmental context of disruptive behaviors in preschool-aged children (Egger & Angold, 2006; Keenan et al., 2011; Keenan & Wakschlag, 2004; Wakschlag et al., 2007). Several DSM-IV diagnostic criteria for ODD and CD, as well as items in the PSC-17 and BPI, refer to behaviors which

can be developmentally normal in very young children. For example, Wakschlag and colleagues (2007) describe noncompliance, temper loss, and aggression as “normative misbehaviors,” or disruptive behaviors which can be expected of typical preschool-aged children. When used as diagnostic criteria for ODD or CD without accounting for developmental norms, these behaviors may result in the over-diagnosis of disruptive behavior disorders. At the other end of the spectrum, several behavioral symptoms (e.g., stealing using force, truancy) are either so extreme or so improbable that their use as screening items may contribute to under-identification of preschoolers with disruptive behavior disorders (Wakschlag et al., 2007). Specification of the developmental appropriateness of potential items to be used in screening young children for disruptive behavior problems is an important step in identifying an appropriate tool for this purpose.

Despite the potential utility of the PSC-17 and BPI in pediatric primary care, several important questions remain regarding their quality of measurement with preschool-aged children. In particular, the levels of preschoolers’ disruptive behavior problems best measured by items in these instruments are unknown. Knowledge of the amount of measurement precision offered along the spectrum of disruptive behaviors in this population could be invaluable in evaluating and improving the performance of these scales with preschool-aged children; for example, identification of items providing informative and precise measurement at sub-clinical to clinical levels of disruptive behavior problems could facilitate development of a brief screening instrument appropriate for use with very young children in pediatric primary care settings. Similar efforts were reported recently by Gomez (2008), applied to Attention Deficit Hyperactivity Disorder (ADHD) parent- and teacher-rating scales. The current study aimed to overcome the limitations of previous analyses by using item response theory (IRT) to examine the psychometric properties of externalizing subscale items in the PSC-17 and BPI as applied to preschool-aged children, an age range crucial for early identification efforts.

Brief Overview of Item Response Theory

In IRT, patterns of participants’ responses to items are analyzed with respect to the underlying latent construct being measured, commonly referred to as θ (Hambleton & Swaminathan, 1985). In this paper, θ refers to the underlying continuous construct of disruptive behavior problems, ranging from very low levels to very high levels. Products of IRT analyses include parameter estimates describing specific measurement properties of individual items.

Several important advantages are conferred by IRT models. First, the novel data offered by IRT regarding item- and scale-level measurement performance can be generalized from one sample to another after linear transformation, unlike the traditional psychometric indices obtained via traditional Classical Test Theory (CTT; i.e., summed score) methods, which are limited to the samples investigated. Second, several questions which are impossible to address with CTT are answerable with IRT. For example, IRT model-fitting allows comparison of the relative merit of items in terms of the amount of information they provide for measuring specific levels of the underlying construct of interest (Hambleton & Swaminathan, 1985). Information levels can be interpreted as the degree of measurement precision provided by an item at various levels of θ . In the context of evaluating items for use in screening young children for disruptive behavior problems, the assessment and comparison of item information levels allows identification of items which are most informative at “easy” (i.e., low-to-average) versus “difficult” (i.e., sub-clinical to clinical) levels of disruptive behavior problems—with the latter potentially representing the most desirable items for use in screening.

The purpose of this study was to evaluate the psychometric properties of the items measuring disruptive behavior problems in the PSC-17 and BPI when used with preschool-aged children seen in pediatric primary care practices. Item parameter estimates characterized items in terms of their difficulty levels (i.e., the levels of θ measured by their response options, represented by b_{ik}), discrimination levels (i.e., ability to discriminate between respondents at different levels of θ , represented by a_i), and levels of information provided along the continuum of disruptive behavior problems. The goal was to identify items which were most informative at sub-clinical to clinical levels of disruptive behavior problems within the developmental context of preschool-aged children.

Method

Participants

Parents (or other primary caregivers) of preschool-aged children ($N = 900$) were recruited to participate from four sociodemographically diverse, university-affiliated pediatric primary care settings. Two clinics served primarily urban, low socioeconomic status (SES) families. Two served a combination of suburban and rural families with a broad range of SES. Eligible participants were age 18 or older; primary caregivers of at least one child between the ages of 3 and 5 years; and in attendance at pediatric primary care appointments at one of the four clinics. Exclusion criteria included already having responded to the survey regarding another child in the home, and presenting for an emergency appointment.

Of the 938 eligible participants approached in pediatric primary care waiting rooms, 900 parents of children between the ages of 3 and 5 years agreed to participate, yielding a 96% response rate. Approximately equal numbers were recruited from each site. Reasons reported for visits included well child check-ups (26%), sick visits (33%), siblings' appointments (28%), and others (13%), including a wide range of issues from allergy shots to minor injuries to dental care.

Parent ages ranged from 18 to 78 years with a mean of 31 years ($SD = 8$ years). The majority of parents ($n = 776$, 87%) were female. Most identified themselves as either white ($n = 491$, 55%) or African-American ($n = 375$, 42%), with only 3% ($n = 32$) identifying other racial or ethnic backgrounds. Participants were not found to differ significantly from non-responders by sex, race, or clinic. Most ($n = 786$, 88%) identified themselves as biological parents of the target children.

Parents provided demographic and behavioral health information about the target children, summarized in Table 1. Regarding child behavioral health, the prevalence of parental concern regarding child behavior problems ($n = 232$, 26%) and reported receipt of mental health services ($n = 85$, 10%) were comparable to rates reported in previous studies (Keenan & Wakschlag, 2004; Lavigne et al., 1998; Lavigne et al., 2009). Parent characteristics are provided in Table 2.

Procedure

All study procedures were approved by the University of Louisville Institutional Review Board (IRB). Eligible parents of preschool-aged children from each clinic were recruited on various times and days of the week over the course of 8 months of the school year. During each recruitment window, all parents in the waiting areas of each clinic were approached to determine study eligibility and request participation. Following informed consent procedures, parents completed the survey in quiet areas of the clinic waiting rooms. Any parent with more than one child between the ages of 3 and 5 was asked to select the child with the most recent birthday as the target child. Participating parents were entered in a drawing to win one of five \$100 gift cards at the conclusion of the study.

Measures

The study survey included three instruments: a sociodemographic questionnaire and two parent-report scales measuring child behavior problems. The order of the behavior rating scales was counterbalanced in the distributed surveys to avoid response set or order bias.

Sociodemographic questionnaire—Parent- and child-level data were obtained, including age, sex, race, level of household income, years of education completed, relationship to the child, family structure, number of siblings in the home, type of health insurance, and number of hours per week spent in daycare and/or school. Reason for the child's appointment and the child's history of behavioral concerns and treatment were also assessed.

Pediatric Symptom Checklist-17 (PSC-17; Gardner et al., 1999)—This brief version of the Pediatric Symptom Checklist (Jellinek, Murphy, & Burns, 1986) was developed for use in pediatric clinics to screen children for psychosocial problems. Parents rate their child on 17 items using a 3-point Likert-type scale (0 = *never*, 1 = *sometimes*, 2 = *often*). Traditional CTT-based scoring involves summing item responses for a total score, where higher scores indicate higher levels of dysfunction. Possible scores on the entire instrument range from 0 to 34.

Investigations of the factor structure of the PSC-17 suggested that the instrument can be separated into three subscales, including an externalizing subscale (7 items), an internalizing subscale (5 items), and an attention subscale (5 items; Gardner et al., 1999). A cut-score of 7 (range: 0 to 14) on the externalizing subscale indicates the need for further assessment for disruptive behavior problems (Gardner et al., 1999). Psychometric properties of the PSC-17 reported by its authors included high levels of internal consistency for the full scale (Cronbach's $\alpha = .89$), as well as for the externalizing subscale (Cronbach's $\alpha = .83$). When used to identify children with disruptive behavior problems, the externalizing subscale reportedly exhibited a sensitivity of 77% and specificity of 80%, as compared to classifications of problems yielded by the parent-completed Iowa-Connors aggression subscale (Loney & Milich, 1982), a modification of the Conner's Teacher Rating Scale (Conners, 1969) with an author-reported internal consistency reliability coefficient of .86.

Behavior Problems Index (BPI)—The BPI (Peterson & Zill, 1986; Zill, 1990) was developed for use in national longitudinal surveys to measure behavioral problems in children and was standardized on a random sample of 6,000 children (Baker, Keck, Mott, & Quinlan, 1993). Its items were derived from the Child Behavior Checklist (Achenbach & Edelbrock, 1981) to provide a shorter scale appropriate for use in survey research. Parents rate their child on 28 items (26 for preschool-aged children) using a 3-point Likert-type scale (0 = *not true*, 1 = *sometimes true*, 2 = *often true*). Total scores are computed via traditional CTT-based methods, by summing item responses. Higher scores indicate higher levels of dysfunction. Possible scores on the entire instrument range from 0 to 52 for preschool-aged children.

The BPI has six subscales, measuring headstrong behaviors, antisocial behaviors, peer problems, anxious/depressed mood, hyperactivity, and immature dependency (Zill, 1990). Three of these subscales are relevant to the measurement of disruptive behavior problems: the headstrong subscale (5 items), the antisocial subscale (4 items), and the peer problems subscale (2 of 3 items are relevant to disruptive behaviors). For the current study, these three subscales (minus 1 internalizing peer problems item) were combined into a BPI externalizing subscale consisting of 11 items. This measure of disruptive behavior problems is similar to a 15 item measure developed from the BPI by Cooksey, Menaghan, and

Jekielek (1997), excluding 2 items targeting impulsive and inattentive behaviors (associated primarily with ADHD) and 2 items measuring school behavior (not included in the preschool version of the BPI). Possible scores on the BPI externalizing subscale range from 0 to 22. Psychometric properties of the BPI reported in previous studies included high estimates of internal consistency for the full instrument (Cronbach's α ranging from .89 to .90; Gortmaker, Walker, Weitzman, & Sobol, 1990; Zill, 1990), and somewhat lower estimates for individual subscales (Cronbach's α ranging from .63 to .75; Gortmaker et al., 1990; Spencer, Fitch, Grogan-Kaylor, & McBeath, 2005).

Statistical Analyses

Descriptive analyses were conducted using IBM SPSS for Windows, Version 20.0. As is common in item calibration studies intended to inform scale development, the 7 externalizing subscale items of the PSC-17 and the 11 externalizing subscale items of the BPI were combined into a larger item pool so that patterns of responses to all items could be considered in IRT analyses. The 18 items in the item pool are referred to henceforth as the *combined externalizing subscale*. Steps in the IRT analyses included (a) testing IRT model assumptions, and (b) fitting an IRT model to the data to obtain item parameter estimates, item information functions, and subscale information functions.

Three primary assumptions of IRT were tested: unidimensionality, local independence, and monotonicity. Exploratory factor analysis (EFA) was conducted on the 18 items, forcing a single factor with principal axis factoring as the extraction method. Dominance of the first factor was reviewed to assess the dimensionality of the combined externalizing subscale. The assumption of local independence requires that once the level of the underlying construct (i.e., disruptive behavior problems) is controlled, item responses should be statistically independent from one another (Steinberg & Thissen, 1996; Wainer & Thissen, 1996; Yen, 1993). Assessment of local independence involved examination of the residual correlation matrix from the exploratory factor analysis. The assumption of monotonicity for items with ordinal response options refers to the requirement that the probability of selecting progressively higher item response options increases with higher levels of the underlying construct, and never decreases. To assess this assumption, trace lines were generated by fitting a non-parametric IRT model to the data from the combined subscales, and graphical results were visually inspected for the expected form. Evaluations of unidimensionality and local independence were conducted using IBM SPSS for Windows, Version 20.0 (2012), while assessment of trace line functions was achieved using TestGraf software (Ramsay, 2000).

Next, Samejima's (1969) two-parameter graded response model (GRM) was fit to the observed data for the combined externalizing subscale to obtain item parameter estimates using MULTILOG 7.03 software (Thissen, Chen, & Bock, 2003). Model fit was evaluated using default settings of the MODFIT computer program (Stark, 2002), which generated fit plots for each item as well as χ^2 tests described by Drasgow, Levine, Tsien, Williams, & Mead (1995). Finally, to identify the most clinically informative items, parameter estimates for each item were reviewed to identify those exhibiting information peaks at or above +1.5 SD, the cut-point selected *a priori* to operationalize sub-clinical to clinical levels of disruptive behavior problems. This conservative decision rule was intended to align conceptually with the "at risk" cut-points used in broadband behavioral rating scales (e.g., Achenbach & Edelbrock, 1981), while controlling the potential rate of false positives.

Results

Results of EFA demonstrated that a single factor (eigenvalue = 6.53) accounted for 36% of the variance. This exceeded the minimum standard of 20% suggested by Reckase (1979) as

sufficient for a scale to be “unidimensional enough” for IRT analyses. Magnitudes of eigenvalues for additional factors and strength of factor loadings, in combination with visual evaluation of a scree plot, also suggested unidimensionality of the combined externalizing subscale. After extraction of the single factor, absolute values of residual correlations for each pair of items ranged from .00 to .15. Using Reeve and colleagues’ (2007) criterion of $|r| \leq .20$ for violations of local independence, this assumption was adequately met as well. Finally, non-parametric trace lines for each item were visually inspected for the expected form under the GRM. All items clearly exhibited the expected form.

Examination of fit plots for each item suggested overall good fit, though several items displayed some degree of misfit in the tails of one or more option characteristic curves (OCCs; i.e., plots of the probability of each response option to an item along the continuum of the latent construct, θ). Results of the statistical tests of model fit described by Drasgow and colleagues (1995) suggested that the fit of the GRM to the data was acceptable.

Item Parameter Estimates

Each item was characterized by three parameter estimates: a (discrimination), b_1 (difficulty threshold between option 0 and option 1), and b_2 (difficulty threshold between option 1 and option 2). High values of a indicate highly discriminating items, meaning that items are better able to distinguish between participants at similar levels of disruptive behavior problems, as compared to items with lower values of a . Values of the parameters b_1 and b_2 provide the difficulty level of the item via the locations of the intersections of the OCCs along the continuum of disruptive behavior problems. Item parameter estimates and standard errors for each item in the combined externalizing subscale are presented in Table 3, along with traditional psychometric descriptive information regarding item means and corrected item-total correlations.¹ In addition, plots of OCCs for all 18 combined externalizing subscale items are provided in Figure 1, illustrating the meaning of the estimated item parameters.

Discrimination parameters among the 18 calibrated items varied ($M = 1.62$, $SD = 0.34$). The highest quartile of discrimination parameter estimates included those for items assessing having a very strong temper (BPI 19); having trouble getting along with others (BPI 12); not listening to rules (PSC-17 12); and bullying/being cruel or mean (BPI 9). These items exhibited very high precision in measuring certain levels of disruptive behavior problems. The lowest quartile of discrimination parameter estimates were for items assessing being high strung (BPI 3); not understanding others’ feelings (PSC-17 5); cheating/lying (BPI 4); and refusing to share (PSC-17 4). The effects of higher versus lower discrimination parameters can be seen in Figure 1 by comparing the OCC plots for the items tapping bullying/being cruel or mean (BPI 9, part [k]) and being high strung (BPI 3, part [h]), in which the item with the highest discrimination parameter estimate (BPI 9) exhibits steeper curves than the item with the lowest discrimination parameter estimate (BPI 3). The behaviors described in items with low discrimination parameters (e.g., being high strung [BPI 3]) did not distinguish between preschool-aged children at similar levels of disruptive behavior problems as well as those with high discrimination parameters (e.g., bullying/being cruel or mean [BPI 9]).

Difficulty parameter estimates among items differed as well. The distribution of the b_1 difficulty parameter was centered just below the mean level of disruptive behavior problems ($M = -0.17$, $SD = 0.74$). This suggests that the threshold level of disruptive behavior

¹Differential item functioning analyses revealed no significant differences in item parameter estimates by questionnaire ordering condition (results available from the author upon request).

problems required for a randomly selected participant to select response option 1 (*sometimes* or *sometimes true*) rather than response option 0 (*never* or *not true*) was, on average, just below the mean level of disruptive behavior problems. The lowest b_1 parameter estimate was for the item assessing not listening to rules (PSC-17 12), making this item the *easiest* of the set—in other words, very low levels of disruptive behavior problems were necessary for a parent to respond that the child *sometimes* does not follow rules, versus responding *never* to this item. Other items with low b_1 parameter estimates included items tapping refusing to share (PSC-17 4); being disobedient at home (BPI 10); and fighting (PSC-17 8). In contrast, several items exhibited much higher difficulty levels for their lower thresholds: items assessing breaking/destroying things (BPI 22); being high strung (BPI 3); and being disliked by others (BPI 15) had the highest b_1 parameter estimates.

Estimates for the upper difficulty threshold parameter b_2 were also disparate. The distribution of the b_2 difficulty parameter estimates clustered between 1.5 and 2 standard deviations above the mean ($M = 1.91$, $SD = 0.52$). Thus, the average threshold level of disruptive behavior problems required for a randomly selected participant to select response option 2 (*often* or *often true*) rather than response option 1 (*sometimes* or *sometimes true*) was in the sub-clinical to clinical range of disruptive behavior problems. The highest b_2 parameter estimate was for the item assessing being disliked by others (BPI 15), making this item the most *difficult* of the set: Extremely high levels of disruptive behavior problems were necessary for a parent to respond that the child *often* is not liked by other children, versus responding *sometimes* to this item. The other items comprising the highest quartile of b_2 parameter estimates included items tapping having trouble getting along with others (BPI 12); not understanding others' feelings (PSC-17 5); being high strung (BPI 3); and teasing others (PSC-17 14). Several items, however, exhibited much lower difficulty levels for their upper thresholds: items assessing not listening to rules (PSC-17 12); having a very strong temper (BPI 19); arguing too much (BPI 6); and being stubborn, sullen, or irritable (BPI 18) all had b_2 parameter estimates lower than 1.5 standard deviations above the mean level of disruptive behavior problems.

The effects of lower versus higher b_1 and b_2 parameters on overall item functioning can be seen in Figure 1 by comparing the OCC plots for the “easy” item assessing not listening to rules (PSC-17 12, part [e]) versus the “difficult” item assessing being disliked by others (BPI 15, part [o]). The difficulty parameter estimates for item PSC-17 12 locate its entire set of curves further to the left on the continuum of disruptive behavior problems than is seen in more difficult items' plots. The behaviors described in “easy” items were often endorsed even by parents of children with low levels of disruptive behavior problems, suggesting that they may be developmentally typical of preschool-aged children. In contrast, much higher levels of disruptive behavior problems were required for parents to endorse the behaviors described in “difficult” items, suggesting that these behaviors were outside of the normative range.

Item Information

Table 4 summarizes the item information function values for each of the 18 items in the combined externalizing subscale. Item information values are provided for θ levels from -3 SD to $+3$ SD. Bolded values in the table indicate information levels near the peak values of information for each item; these appear at the approximate levels of θ at which each item was most informative. As reported in Table 4, of the 18 items, 5 were most informative only at levels below the sub-clinical range of disruptive behavior problems (i.e., less than $+1.5$ SD): does not listen to rules (PSC-17 12); argues too much (BPI 6); is disobedient at home (BPI 10); is stubborn, sullen, or irritable (BPI 18); and has a very strong temper (BPI 19). In the context of screening efforts aimed at identifying children at risk of disruptive behavior disorders, these items may not contribute useful measurement information; they may be

more useful in measuring developmentally typical behaviors of preschool-aged children for applications other than screening. The remaining 13 items, however, yielded varying levels of information along the range of sub-clinical to clinical disruptive behavior problems (i.e., greater than +1.5 SD), the atypical levels which would be targeted in screening efforts for young children.

Discussion

Early identification of disruptive behavior problems in preschool-aged children is a key step in the prevention of later behavioral issues. Screening in pediatric primary care with reliable and valid instruments is one promising approach for improving early identification efforts. However, existing screening tools may be inappropriate for use with very young children, a group whose developmental context complicates screening efforts. This study applied IRT analyses to evaluate the psychometric properties of 18 externalizing subscale items from two instruments—the PSC-17 and BPI—with preschool-aged children seen in pediatric primary care practices. The items were characterized in terms of their difficulty levels (b_{ik}), discrimination levels (a_i), and levels of information provided along the continuum of disruptive behavior problems observed in very young children. The goal was to identify items which were most informative at sub-clinical to clinical levels of disruptive behavior problems within the developmental context of preschool-aged children.

The purpose of screening for disruptive behavior problems is to identify children who may benefit from further assessment (followed by treatment, if indicated); thus, desirable items in a parent-report screening tool should target higher levels of disruptive behaviors than typically seen among preschool-aged children. In the current study, a difficulty level of 1.5 SD above the mean was selected *a priori* as the lower bound for clinically concerning disruptive behaviors. Consistent with the developmental framework proposed by Wakschlag and colleagues (2007) for identifying disruptive behaviors in preschool children, items providing measurement information only at lower levels of disruptive behaviors could be considered normative misbehaviors, potentially contributing to over-identification of children if included in a screening tool due to their developmental imprecision.

Results of IRT analyses revealed that of the 18 investigated items, 5 items measured only low levels (less than +1.5 SD) of disruptive behaviors among preschool-aged children. These items' stems described not following rules (PSC-17 12); arguing too much (BPI 6); being disobedient at home (BPI 10); being sullen, stubborn, or irritable (BPI 18); and having a very strong temper (BPI 19). In the context of typical development among children ages 3 to 5, it is not surprising that these items would be characterized by low difficulty levels. As noncompliance, temper loss, and aggression can be described as normative misbehaviors among preschool-aged children (Wakschlag et al., 2007), parents' endorsements of these items likely represent typical behaviors rather than indicators of clinical concern.

The upper difficulty thresholds of the remaining 13 items all exceeded +1.5 SD, indicating that they measure higher levels of disruptive behavior problems. However, 5 of these difficult items offered comparatively less measurement information, suggesting suboptimal measurement in the sub-clinical to clinical range of disruptive behaviors. These less precise items tap behaviors including refusing to share (PSC-17 4), not understanding others' feelings (PSC-17 5), teasing others (PSC-17 14), being high strung, nervous, or tense (BPI 3), and cheating/lying (BPI 4). Items yielding low amounts of measurement information may be poor indicators of behavioral problems in preschool-aged children for several reasons, including variations in typical child development, inappropriateness of item content for preschool-aged children, and differences in item interpretation among parents. Several of

these items (e.g., BPI 3 and BPI 4) are also double-barreled questions, perhaps contributing to measurement imprecision.

The remaining 8 items had high discrimination and difficulty parameters, offering considerable measurement information at sub-clinical to clinical levels of disruptive behavior problems. Behaviors targeted by these items included bullying/cruelty to others (BPI 9), lack of remorse after misbehavior (BPI 11), difficulty getting along with other children (BPI 12), not being liked by other children (BPI 15), deliberately breaking/destroying things (BPI 22), fighting with other children (PSC-17 8), blaming others (PSC-17 10), and taking things that do not belong to him/her (PSC-17 16). Apart from not feeling sorry after misbehavior and peer rejection, each of these behaviors has been shown to distinguish clinically referred from non-referred preschoolers (Keenan & Wakschlag, 2004). In addition, early peer relationship problems may predict CD symptoms in adolescence (Emond, Ormel, Veenstra, & Oldehinkel, 2007), and lack of remorse is a symptom of Antisocial Personality Disorder diagnosed in adolescence or adulthood, associated with childhood-onset CD (Hill, 2003; Lahey, Loeber, Burke, & Applegate, 2005). Thus, these 8 items appear to represent clinically concerning behaviors within the developmental context of preschool-aged children.

The overlap of the content of the 18 combined externalizing subscale items with specific diagnostic criteria for ODD and CD is potentially informative in conceptualizing the spectrum of disruptive behavior problems in preschoolers. Interestingly, the content of the 5 items measuring only low levels of disruptive behaviors referenced four diagnostic criteria for ODD (temper loss, defiance, arguing, and being easily annoyed), but no diagnostic criteria for CD. The 5 items with low discrimination and imprecise measurement of sub-clinical to clinical levels of disruptive behaviors included only one specific diagnostic criterion for either disorder: lying, a symptom of CD. The remaining 4 low-information items address behaviors not specifically included in the diagnostic criteria for either ODD or CD (i.e., not sharing, not understanding others' feelings, teasing others, and being high strung); however, the low-information item addressing not understanding others' feelings may target lack of empathy, a characteristic under the proposed Callous and Unemotional specifier for CD in the upcoming DSM revision (Frick, 2012).

In contrast, the 8 items offering high levels of measurement information at sub-clinical to clinical levels of disruptive behaviors reflected one symptom of ODD (blaming others) and five symptoms of CD (fighting, bullying, cruelty to others, property destruction, and stealing without confrontation), with the remaining items referring to lack of remorse and poor peer relationships, frequently associated with CD (Emond et al., 2007; Miller-Johnson, Coie, Maumary-Gremaud, Bierman, & Conduct Problems Prevention Research Group, 2002). Interestingly, lack of remorse also falls under the proposed Callous and Unemotional specifier for CD. This item may provide clinically important screening information about a type of CD potentially associated with particularly severe problems with aggression and cruelty (Kahn, Frick, Youngstrom, Findling, & Youngstrom, 2011).

No consensus exists regarding whether ODD and CD are two distinct disorders or simply different degrees of one disorder (Egger & Angold, 2006; Loeber, Burke, Lahey, Winters, & Zera, 2000). These results seem consistent with a conceptualization of the two disorders as hierarchical on the same spectrum, with CD symptoms indicating higher levels of disruptive behavior problems than ODD symptoms (Lahey, Loeber, Quay, Frick, & Grimm, 1997).

Limitations

Results should be interpreted in the context of several study limitations. First, recent literature has emphasized the importance of measuring impairment associated with

symptoms of disruptive behavior disorders, not just the presence, absence, or categorical frequency (i.e., “never,” “often”) of behavioral symptoms (Keenan, 2012). This is especially salient in light of anticipated revisions to DSM criteria for disruptive behavior disorders. In the context of screening rather than diagnostic assessment, however, determination of impairment may not be necessary, as some over-identification of children in need of assessment may be preferable to under-identification to achieve the benefits of early identification. Similarly, reliance solely on parent-report is generally inadequate in diagnostic assessment, but is efficient and appropriate for the purposes of brief screening. In addition, only 18 items were assessed, which may not include all disruptive behaviors salient to screening preschool-aged children. The set of items analyzed, however, yielded a wide distribution of difficulty and discrimination parameters and were drawn from previously established instruments. Finally, participants were recruited from a single geographic area, potentially limiting generalizability of findings.

Conclusions and Future Directions

Results of IRT analyses investigating 18 potential items for screening preschool-aged children for disruptive behavior problems revealed 8 items which were highly informative at sub-clinical to clinical levels of disruptive behaviors. Behaviors measured by these items were consistent with those identified in recent efforts to distinguish developmentally typical misbehaviors from clinically concerning behaviors among preschool-aged children (Keenan & Wakschlag, 2004; Wakschlag et al., 2007). Further investigation of the performance of these items with target children of diverse sociodemographic backgrounds may facilitate the selection of those most appropriate for inclusion in a screening tool designed for pediatric primary care. Specifically, these items should be tested for differential item functioning (DIF) when used with children in differing sociodemographic groups (e.g., by sex and race). Unbiased items which are highly informative at sub-clinical to clinical levels of disruptive behavior problems should be assembled into a very brief screening tool and assessed for validity in identifying children with diagnoses of disruptive behavior disorders (i.e., ODD and CD). Sensitivity, specificity, positive predictive value, negative predictive value, and test-retest reliability of such a tool should be determined. Once validated, a new parent-report screening tool incorporating these items would be efficient and appropriate for the fast-paced primary care setting, where early identification of disruptive behavior problems is a pressing need.

Acknowledgments

This manuscript was completed with the support of an award (8KL2TR000116-02) from the University of Kentucky CTSA (UL1RR033173), supported by the National Center for Research Resources, now at the National Center for Advancing Translational Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

The authors wish to thank the physicians, staff, and patients of the University of Louisville Department of Pediatrics and Oldham County Pediatrics, as well as Gerard M. Barber, Ph.D., James Clark, Ph.D., Andrew J. Frey, Ph.D., and V. Faye Jones, M.D., for their support. The authors also acknowledge the invaluable data collection assistance of Judith Friedrich, Ph.D., Demeka Campbell, M.D., and Cynthia Bowman-Stroud, M.D.

References

- Achenbach TM, Edelbrock C. Behavioral problems and competencies reported by parents of normal and disturbed children aged four through sixteen. Monographs of the Society for Research in Child Development. 1981; 46:1–82. [PubMed: 7242540]
- Agency for Healthcare Research and Quality. Guide to clinical preventive service, 3rd edition: Systematic evidence reviews. Rockville, MD: Agency for Healthcare Research and Quality; 2002.

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-IV (4th ed.). Washington, DC: American Psychiatric Association; 1996.
- Arroll B, Goodyear-Smith F, Kerse N, Fishman T, Gunn J. Effect of the addition of a 'help' question to two screening questions on specificity for diagnosis of depression in general practice: Diagnostic validity study. *British Medical Journal*. 2005; 331:884–887. [PubMed: 16166106]
- Baker, PC.; Keck, CK.; Mott, FL.; Quinlan, SV. National longitudinal survey of youth handbook (revised edition). Columbus, OH: The Ohio State University, Center for Human Resource Research; 1993.
- Berle D, Starcevic V, Moses K, Hannan A, Milicevic D, Sammut P. Preliminary validation of an ultra-brief version of the Penn State Worry Questionnaire. *Clinical Psychology and Psychotherapy*. 2011; 18:339–346. [PubMed: 20806422]
- Bradley KA, DeBenedetti AF, Volk RJ, Williams EC, Frank D, Kivlahan DR. AUDIT-C as a brief screen for alcohol use in primary care. *Alcoholism: Clinical and Experimental Research*. 2007; 31:1208–1217.
- Conners CK. A teacher rating scale for use in drug studies with children. *American Journal of Psychiatry*. 1969; 126:884–888. [PubMed: 4900822]
- Cooksey EC, Menaghan EG, Jekielek SM. Life-course effects of work and family circumstances on children. *Social Forces*. 1997; 76:637–665.
- Costello EJ, Edelbrock CS. Detection of psychiatric disorders in pediatric primary care: A preliminary report. *Journal of the American Academy of Child Psychiatry*. 1985; 24:771–774. [PubMed: 4067146]
- Costello EJ, Shugart MA. Above and below the threshold: Severity of psychiatric symptoms and functional impairment in a pediatric sample. *Pediatrics*. 1992; 90:359–368. [PubMed: 1518689]
- Dragow F, Levine MV, Tsien S, Williams B, Mead AD. Fitting polytomous item response theory models to multiple-choice tests. *Applied Psychological Measurement*. 1995; 19:143–165.
- Egger HL, Angold A. Common emotional and behavioral disorders in preschool children: Presentation, nosology, and epidemiology. *Journal of Child Psychology and Psychiatry*. 2006; 47:313–337. [PubMed: 16492262]
- Emond A, Ormel J, Veenstra R, Oldehinkel AJ. Preschool behavioral and social-cognitive problems as predictors of (pre)adolescent disruptive behavior. *Child Psychiatry & Human Development*. 2007; 38:221–236. [PubMed: 17476586]
- Fanton JH, MacDonald B, Harvey EA. Preschool parent-pediatrician consultations and predictive referral patterns for problematic behaviors. *Journal of Developmental and Behavioral Pediatrics*. 2008; 29:475–482. [PubMed: 18941427]
- Frick PJ. Developmental pathways to conduct disorder: Implications for future directions in research, assessment, and treatment. *Journal of Clinical Child & Adolescent Psychology*. 2012; 41:378–389. [PubMed: 22475202]
- Gardner W, Murphy M, Childs G, Kelleher KJ, Pagano M, Jellinek MS, Chiapetta L. The PSC-17: A brief Pediatric Symptom Checklist including psychosocial problem subscales. A report from PROS and ASPN. *Ambulatory Child Health*. 1999; 5:225–236.
- Gomez R. Item response theory analyses of the parent and teacher ratings of the DSM-IV ADHD rating scale. *Journal of Abnormal Child Psychology*. 2008; 36:865–885. [PubMed: 18266103]
- Gortmaker SL, Walker DK, Weitzman M, Sobol AM. Chronic conditions, socioeconomic risks, and behavioral problems in children and adolescents. *Pediatrics*. 1990; 85:267–276. [PubMed: 2304779]
- Halfon, N.; Regalado, M.; McLearn, KT.; Kuo, AA.; Wright, K. Building a bridge from birth to school: Improving developmental and behavioral health services for young children. New York: The Commonwealth Fund; 2003. publication no. 564.
- Hambleton, RK.; Swaminathan, H. Item response theory: Principles and applications. Norwell, MA: Kluwer Academic Publishers; 1985.
- Harvey EA, Youngwirth SD, Thakar DA, Errazuriz PA. Predicting attention-deficit/hyperactivity disorder and oppositional defiant disorder from preschool diagnostic assessments. *Journal of Consulting and Clinical Psychology*. 2009; 77:349–354. [PubMed: 19309194]

- Hill J. Early identification of individuals at risk for antisocial personality disorder. *British Journal of Psychiatry*. 2003; 182(Suppl. 144):11–14.
- Hill LG, Lochman JE, Coie JD, Greenberg MT. Effectiveness of early screening for externalizing problems: Issues of screening accuracy and utility. *Journal of Consulting and Clinical Psychology*. 2004; 72:809–820. [PubMed: 15482039]
- Hinkin TR. A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*. 1998; 1:104–121.
- Jellinek MS, Murphy JM, Burns BJ. Brief psychosocial screening in outpatient pediatric practice. *Journal of Pediatrics*. 1986; 109:371–378. [PubMed: 3734977]
- Kahn RE, Frick PJ, Youngstrom E, Findling RL, Youngstrom JK. The effects of including a callous-unemotional specifier for the diagnosis of conduct disorder. *Journal of Child Psychology and Psychiatry*. 2011; 53:271–282. [PubMed: 21950481]
- Kataoka SH, Zhang L, Wells KB. Unmet need for mental health care among U.S. children: Variation by ethnicity and insurance status. *American Journal of Psychiatry*. 2002; 159:1548–1555. [PubMed: 12202276]
- Keenan K. Mind the gap: Assessing impairment among children affected by proposed revisions to the diagnostic criteria for oppositional defiant disorder. *Journal of Abnormal Psychology*. 2012; 121:352–359. [PubMed: 21707124]
- Keenan K, Boeldt D, Chen D, Coyne C, Donald R, Duax J, Humphries M. Predictive validity of DSM-IV oppositional defiant and conduct disorders in clinically referred preschoolers. *Journal of Child Psychology and Psychiatry*. 2011; 52:47–55. [PubMed: 20738448]
- Keenan K, Shaw DS, Walsh B, Delliquadri E, Giovannelli J. DSM-III-R disorders in preschool children from low-income families. *Journal of the American Academy of Child and Adolescent Psychiatry*. 1997; 36:620–627. [PubMed: 9136496]
- Keenan K, Wakschlag LS. Are oppositional defiant and conduct disorder symptoms normative behaviors in preschoolers? A comparison of referred and nonreferred children. *American Journal of Psychiatry*. 2004; 161:356–358. [PubMed: 14754786]
- Kim-Cohen J, Arseneault L, Caspi A, Tomas MP, Taylor A, Moffitt TE. Validity of DSM-IV conduct disorder in 41/2–5-year-old children: A longitudinal epidemiological study. *American Journal of Psychiatry*. 2005; 162:1108–1117. [PubMed: 15930059]
- Lahey BB, Loeber R, Burke JD, Applegate B. Predicting future antisocial personality disorder in males from a clinical assessment in childhood. *Journal of Consulting and Clinical Psychology*. 2005; 73:389–399. [PubMed: 15982137]
- Lahey, BB.; Loeber, R.; Quay, HC.; Frick, PJ.; Grimm, J. Oppositional defiant disorder and conduct disorder. In: Widiger, TA.; Frances, AJ.; Pincus, HA.; Ross, R.; First, MB.; Davis, W., editors. *DSM-IV Sourcebook*. Vol. Vol. 3. Washington DC: American Psychiatric Association; 1997. p. 189-209.
- Lavigne JV, Arend R, Rosenbaum D, Binns HJ, Christoffel KK, Burns A, Smith A. Mental health service use among young children receiving pediatric primary care. *Journal of the American Academy of Child and Adolescent Psychiatry*. 1998; 37:1175–1183. [PubMed: 9808929]
- Lavigne JV, Binns HJ, Christoffel KK, Rosenbaum D, Arend R, Smith K, McGuire PA. Behavioral and emotional problems among preschool children in pediatric primary care: Prevalence and pediatricians' recognition. *Pediatrics*. 1993; 91:649–655. [PubMed: 8441575]
- Lavigne JV, Lebailly SA, Hopkins J, Gouze KR, Binns HJ. The prevalence of ADHD, ODD, depression, and anxiety in a community sample of 4-year-olds. *Journal of Clinical Child and Adolescent Psychology*. 2009; 38:315–328. [PubMed: 19437293]
- Loeber R, Burke JD, Lahey BB, Winters A, Zera M. Oppositional defiant and conduct disorder: A review of the past 10 years, part I. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2000; 39:1468–1484.
- Loney, J.; Milich, R. Hyperactivity, inattention, and aggression in clinical practice. In: Wolraich, ML.; Routh, D., editors. *Advances in developmental and behavioral pediatrics*. Vol. Vol. 3. Greenwich, CT: JAL; 1982. p. 113-147.

- Miller-Johnson S, Coie JD, Maumary-Gremaud A, Bierman K, Conduct Problems Prevention Research Group. Peer rejection and aggression and early starter models of conduct disorder. *Journal of Abnormal Child Psychology*. 2002; 30:217–230. [PubMed: 12041708]
- Peterson JL, Zill N. Marital disruption, parent-child relationships, and behavior problems in children. *Journal of Marriage and the Family*. 1986; 48:295–307.
- Pierce EW, Ewing LJ, Campbell SB. Diagnostic status and symptomatic behavior of hard-to-manage preschool children in middle childhood and early adolescence. *Journal of Clinical Child Psychology*. 1999; 28:44–57. [PubMed: 10070606]
- Ramsay, JO. TestGraf: A program for the graphical analysis of multiple choice test and questionnaire data. Montreal: McGill University; 2000.
- Reckase MD. Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*. 1979; 4:207–230.
- Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA. PROMIS Cooperative Group (2007). Psychometric evaluation and calibration of Health-Related Quality of Life item banks plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical Care*. 45(5 Suppl 1):S22–S31. [PubMed: 17443115]
- Samejima F. Calibration of latent ability using a response pattern of graded scores. *Psychometrika Monograph*. 1969; (Supplement, 17)
- Spencer MS, Fitch D, Grogan-Kaylor A, McBeath B. The equivalence of the Behavior Problem Index across U.S. ethnic groups. *Journal of Cross-Cultural Psychology*. 2005; 36:573–589.
- Stark, S. MODFIT [computer program]. Urbana Champaign, IL: University of Illinois IRT Modeling Lab; 2002.
- Steinberg L, Thissen D. Uses of item response theory and the testlet concept in the measurement of psychopathology. *Psychological Methods*. 1996; 1:81–97.
- Tabachnick, BG.; Fidell, LS. Using multivariate statistics (4th ed.). Needham Heights, MA: Allyn and Bacon; 2001.
- Thissen, D.; Chen, W-H.; Bock, RD. MULTILOG 7.03 [computer program]. Lincolnwood, IL: Scientific Software International; 2003.
- Wainer H, Thissen D. How is reliability related to the quality of test scores? What is the effect of local dependence on reliability? *Educational Measurement: Issues and Practice*. 1996; 15:22–29.
- Wakschlag LS, Briggs-Gowan MJ, Carter AS, Hill C, Danis B, Keenan K, Leventhal BL. A developmental framework for distinguishing disruptive behavior from normative misbehavior in preschool children. *Journal of Child Psychology and Psychiatry*. 2007; 48:976–987. [PubMed: 17914998]
- Wilson, JMG.; Jungner, G. Principles and practice of screening for disease. Geneva: World Health Organization; 1968. Public Health Papers #34. from http://whqlibdoc.who.int/php/WHO_PHP_34.pdf [Accessed on January 16, 2013]
- Yen W. Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*. 1993; 30:187–213.
- Zill, N. Behavior problem index based on parent report. Washington, DC: Child Trends; 1990.

(a) PSC-17 Item 4

(d) PSC-17 Item 10

(e) PSC-17 Item 16

(j) BPI Item 6

(b) PSC-17 Item 5

(e) PSC-17 Item 12

(h) BPI Item 3

(k) BPI Item 9

(c) PSC-17 Item 8

(f) PSC-17 Item 14

(i) BPI Item 4

(l) BPI Item 10

(m) BPI Item 11

(p) BPI Item 18

(n) BPI Item 12

(q) BPI Item 19

(o) BPI Item 15

(r) BPI Item 22

Figure 1.

Plots of graded response model option characteristic curves (OCCs) for all PSC-17 and BPI items measuring disruptive behavior problems (i.e., the combined externalizing subscale)

Table 1

Target Child Characteristics (N = 900)

Variable	Frequency	(%)
Sex		
Male	472	(53)
Female	424	(47)
Race		
White	450	(50)
African-American	362	(40)
Other	88	(10)
Household Composition		
Two-parent	512	(57)
Single parent	339	(38)
Caregiver other than parent	47	(5)
Health Insurance		
Public	634	(71)
Private	252	(28)
None	10	(1)
Parent believes child has behavior problems	232	(26)
Child has seen a mental health provider	85	(10)
Child has been prescribed medication(s) for behavior	42	(5)

Note. Percentages may not sum to 100 percent due to rounding.

Table 2

Parent Characteristics (N = 900)

Variable	Frequency	(%)
Sex		
Male	118	(13)
Female	776	(87)
Race		
White	491	(55)
African-American	375	(42)
Other	32	(3)
Household Income		
< \$10,000	248	(28)
\$10,001 – \$20,000	187	(21)
\$20,001 – \$40,000	224	(25)
\$40,001 – \$60,000	104	(12)
\$60,001 – \$80,000	45	(5)
> \$80,000	78	(9)
Education		
Less than high school	145	(16)
High school diploma/GED	388	(44)
More than high school	355	(40)

Note. Percentages may not sum to 100 percent due to rounding.

Table 3

Item Descriptives and Graded Response Model Parameter Estimates (N = 900)

Item Descriptives		Parameter Estimates				
Item	Short Wording	M (SD)	r _{ii}	a _i (se)	b _{1i} (se)	b _{2i} (se)
PSC-17 4	Refuses to share	0.85 (0.60)	.47	1.29 (0.12)	-1.12 (0.11)	1.89 (0.17)
PSC-17 5	Does not understand others' feelings	0.66 (0.62)	.44	1.21 (0.11)	-0.43 (0.09)	2.37 (0.23)
PSC-17 8	Fights others	0.81 (0.59)	.58	1.94 (0.15)	-0.82 (0.07)	1.65 (0.12)
PSC-17 10	Blames others	0.59 (0.66)	.52	1.47 (0.13)	-0.07 (0.07)	1.89 (0.16)
PSC-17 12	Does not listen to rules	1.00 (0.59)	.61	2.07 (0.16)	-1.28 (0.09)	1.11 (0.09)
PSC-17 14	Teases others	0.50 (0.60)	.49	1.34 (0.13)	0.11 (0.08)	2.53 (0.22)
PSC-17 16	Takes things	0.65 (0.64)	.53	1.50 (0.13)	-0.33 (0.07)	1.92 (0.16)
BPI 3	High strung	0.43 (0.65)	.43	1.10 (0.12)	0.64 (0.11)	2.49 (0.27)
BPI 4	Cheats/lies	0.67 (0.65)	.49	1.26 (0.11)	-0.37 (0.09)	2.08 (0.19)
BPI 6	Argues too much	0.79 (0.72)	.55	1.43 (0.13)	-0.54 (0.08)	1.37 (0.13)
BPI 9	Bullies/cruel or mean	0.45 (0.63)	.64	2.27 (0.19)	0.31 (0.06)	1.73 (0.12)
BPI 10	Disobedient at home	0.86 (0.62)	.56	1.72 (0.15)	-0.92 (0.08)	1.52 (0.12)
BPI 11	Not sorry after misbehaves	0.49 (0.65)	.55	1.61 (0.14)	0.24 (0.07)	1.96 (0.16)
BPI 12	Trouble getting along with others	0.38 (0.57)	.60	2.02 (0.17)	0.45 (0.06)	2.22 (0.17)
BPI 15	Not liked by others	0.14 (0.39)	.44	1.65 (0.21)	1.65 (0.15)	3.17 (0.37)
BPI 18	Stubborn, sullen, or irritable	0.87 (0.67)	.52	1.41 (0.12)	-0.92 (0.10)	1.43 (0.13)
BPI 19	Very strong temper	0.70 (0.73)	.63	1.99 (0.16)	-0.22 (0.06)	1.21 (0.09)
BPI 22	Breaks/destroys things	0.37 (0.62)	.58	1.88 (0.17)	0.61 (0.07)	1.91 (0.14)

Note. r_{ij} = corrected item-total correlation; a_i = item discrimination parameter; se = standard error; b_{1i} = item lower threshold difficulty parameter; b_{2i} = item upper threshold difficulty parameter; PSC-17 = Pediatric Symptom Checklist-17 (Gardner et al., 1999); BPI = Behavior Problem Index (Zill, 1990).

Table 4

Item Information Function Values for Combined Externalizing Scale Items along the Continuum of Disruptive Behavior Problems (N = 900)

Item	Short Wording	Level of θ^a						
		-3.0	-2.0	-1.0	0.0	+1.0	+2.0	+3.0
PSC-17 4	Refuses to share	0.12	0.31	0.42	0.33	0.36	0.42	0.26
PSC-17 5	Does not understand others' feelings	0.06	0.17	0.33	0.37	0.32	0.37	0.32
PSC-17 8	Fights others	0.05	0.31	0.92	0.63	0.71	0.84	0.24
PSC-17 10	Blames others	0.03	0.11	0.35	0.57	0.54	0.55	0.30
PSC-17 12	Does not listen to rules	0.12	0.64	1.01	0.57	1.07	0.51	0.08
PSC-17 14	Teases others	0.03	0.09	0.27	0.46	0.42	0.45	0.41
PSC-17 16	Takes things	0.04	0.16	0.44	0.57	0.51	0.58	0.31
BPI 3	High strung	0.02	0.06	0.15	0.27	0.34	0.34	0.29
BPI 4	Cheats/lies	0.05	0.16	0.34	0.41	0.38	0.41	0.29
BPI 6	Argues too much	0.06	0.20	0.47	0.54	0.55	0.43	0.17
BPI 9	Bullies/cruel or mean	0.00	0.03	0.24	1.15	1.19	1.18	0.26
BPI 10	Disobedient at home	0.08	0.34	0.74	0.54	0.66	0.63	0.20
BPI 11	Not sorry after misbehaves	0.02	0.07	0.27	0.64	0.66	0.67	0.34
BPI 12	Trouble getting along with others	0.00	0.03	0.19	0.84	0.93	1.03	0.58
BPI 15	Not liked by others	0.00	0.01	0.03	0.16	0.52	0.74	0.74
BPI 18	Stubborn, sullen, or irritable	0.09	0.29	0.51	0.46	0.50	0.43	0.18
BPI 19	Very strong temper	0.15	0.11	0.57	1.05	1.05	0.57	0.11
BPI 22	Breaks/destroys things	0.00	0.03	0.15	0.65	0.96	0.92	0.36

Note. Bolded values indicate levels of θ near which the item information functions peak for each item (actual peaks may be between shown levels of θ). PSC-17 = Pediatric Symptom Checklist-17 (Gardner et al., 1999); BPI = Behavior Problem Index Zill, 1990).

^aRepresents continuum of disruptive behavior problems, with an arbitrary mean of 0 and standard deviation of 1. Column headings refer to standard deviations from the mean.