Published in final edited form as:
Acad Pediatr. 2012 ; 12(5): 456–467. doi:10.1016/j.acap.2012.06.008.

The Preschool Pediatric Symptom Checklist (PPSC): Development and initial validation of a new social/emotional screening instrument

R. Christopher Sheldrick, Ph.D.,
Developmental-Behavioral Pediatrics, Floating Hospital for Children Tufts Medical Center, Boston, MA

Brandi S. Henson, Psy.D.,
Developmental-Behavioral Pediatrics, Floating Hospital for Children Tufts Medical Center, Boston, MA

Shela Merchant, M.A.,
Developmental-Behavioral Pediatrics, Floating Hospital for Children Tufts Medical Center, Boston, MA

Emily N. Neger, B.A.,
Developmental-Behavioral Pediatrics, Floating Hospital for Children Tufts Medical Center, Boston, MA

J. Michael Murphy, Ed.D., and
Department of Psychiatry, Massachusetts General Hospital, Boston, MA

Ellen C. Perrin, M.D.
Developmental-Behavioral Pediatrics, Floating Hospital for Children, Tufts Medical Center, Boston, MA

Abstract

Objective—This paper describes the development and initial validation of the Preschool Pediatric Symptom Checklist (PPSC), a social/emotional screening instrument for children 18–60 months of age. The PPSC was created as one part of a comprehensive screening instrument designed for pediatric primary care and is modeled after the Pediatric Symptom Checklist.

Method—Items for the PPSC were developed by a team of experts who reviewed existing assessment instruments and relevant research literature. Scale construction and initial validation (including factor analysis and tests of construct validity) were conducted with 292 families from pediatric primary care sites and 354 families from referral clinics. 171 additional families were recruited from primary care sites to obtain an independent replication sample.

Results—Exploratory factor analysis revealed 4 dimensions of the PPSC: *Externalizing, Internalizing, Attention Problems*, and *Parenting Challenges.* These dimensions were incorporated

Please address all correspondence to: R. Christopher Sheldrick, Ph.D, 800 Washington Street, Box 854, Boston, MA 02111, Phone: 617-636-4830, Fax: 617-636-8035, rsheldrick@tuftsmedicalcenter.org.

There are no potential conflicts of interest or corporate sponsors to report.

into a bifactor model that displayed a strong general factor, thus supporting the use of a total score. The PPSC total score shows strong internal and retest reliability, and it identifies children who score in the clinical range of a longer, well-validated and more comprehensive parent-report instrument (the CBCL), as well as children who are reported to have a range of behavioral diagnoses. Moreover, sensitivity and specificity with respect to these criteria were comparable to those of another well-accepted but longer screener, the ASQ:SE. Finally, results for the PPSC's total scale remained consistent when replicated in an independent sample.

Conclusion—The PPSC shows promise as a social/emotional screening instrument for use in pediatric primary care.

Keywords

Social; Emotional; Behavioral; Screening; Pediatrics

Introduction

Approximately 10–15 percent of preschool children experience social/emotional problems that cause significant suffering to the child and family[1,2] and may lead to later impairments in quality of life and functional status.[3] Most mental health disorders in adulthood have their roots in childhood,[4,5] suggesting that early identification may reduce long term disability. However, identifying such disabilities in young children can be difficult. Primary care physicians see patients for only a short time and outside their natural settings. Research has shown that when pediatricians use only clinical impressions rather than formal screening, they frequently fail to identify children with developmental-behavioral problems.[6]

Fortunately, parents have extensive knowledge of and can report on their children's behavior. Soliciting parents' observations via validated screening tests allows for comparison to normative data and systematic interpretation, thus providing an efficient way to identify those at risk for social/emotional problems. To be effective in primary care, a screening instrument must be both accurate and feasible—ideally short, easy to read and understand, simple to score, and inexpensive or free to use. While several tools are currently available to identify emotional and behavioral problems in young children, their feasibility may be hampered by length, cost and/or complexity.[7]

One example of a screening tool that meets feasibility criteria is the *Pediatric Symptom Checklist* (PSC). Specifically developed to assess school-aged children in primary care settings, the 35-item PSC was designed with pediatric primary care use in mind. Questions are short and the reading level is low. While scores on three subscales can be calculated, the primary use of the PSC has focused on a single overall scale that can be scored rapidly and compared to a predetermined threshold to ascertain the need for further evaluation. The PSC is freely available and can be downloaded from the Web.[8] The PSC has been well validated across a range of studies,[9–12] and it has become very popular as a screening instrument in pediatrics.[13,14] The Task Force on Mental Health of the American Academy of Pediatrics (AAP) cites the PSC as an example of a mental health screening tool that is appropriate for use in pediatric primary care,[15] and the PSC has been endorsed provisionally as a child health outcome measure by the National Quality Forum.[16] Most validation work using the PSC has focused on the overall psychosocial functioning of 6–12 year old children, [9, 17–19] though some studies have supported its utilization with children as young as 4 years.[20–22]

Because of the importance of early detection of emotional-behavioral problems and the differing developmental and behavioral expectations for infants and preschoolers compared to school-aged children, we set out to create a screening instrument designed to focus more specifically on the social/emotional health of younger children. We worked with the original

developers of the PSC to create two new instruments: the Baby Pediatric Symptom Checklist (BPSC) for children below 18 months, and the Preschool Pediatric Symptom Checklist (PPSC) for children from 18 to 60 months. The creation of these two measures occurred as part of an ongoing project to develop a comprehensive surveillance instrument for children under 5 years of age, known as the *Survey of Wellbeing of Young Children* (SWYC). The SWYC includes four components, which assess social/emotional functioning; cognitive, motor and language development; autism; and family risk factors. The BPSC and PPSC were created to address the social/emotional component of the SWYC.

In this paper we describe the development and initial validation of the PPSC. Like the PSC, the PPSC is designed to maximize feasibility in clinical settings: it is easy to score, freely available and brief enough that it can be administered together with instruments that screen for other problems, such as developmental delays and autism.

Methods

Overview

Based on an extensive review of existing assessment measures and relevant research literature, as well as feedback from parents of young children and experts in child development, we created a list of candidate items for the PPSC. We enrolled two samples of parents to further develop and validate the PPSC: one from primary care sites and one from referral clinics. Using these data, we conducted analyses to reduce the number of items and create a final version of the PPSC, and to obtain estimates of internal reliability and construct validity. To replicate our results, we then enrolled an independent sample from a different set of primary care pediatric practices. To establish retest reliability, approximately 1/3 of this second sample was asked to complete the PPSC a second time 3–4 weeks later. All studies were approved by the Institutional Review Board of Tufts Medical Center.

Item Development and Description

Our goal was to write items that could be easily and efficiently answered in the context of a pediatric waiting room by parents from a range of educational and cultural backgrounds. Thus, we sought to write questions that were short, easy to read, salient to parents and appropriate for children age 18 months through the preschool years. We began by identifying constructs common across several parent-report measures that had previously been validated for children under 5 years, including the PSC, the *Child Behavior Checklist 1.5–5 years* (CBCL),[23] the *Infant-Toddler Social and Emotional Assessment* (ITSEA),[24] and the *Ages & Stages Questionnaire: Social/Emotional* (ASQ:SE).[25] In addition, we reviewed relevant literature and generated items based on our clinical experience. Items were developed to encompass four domains of interest, including three that are included in the PSC (Internalizing, Externalizing, and Attention Problems)[26] and one new domain, Parenting Challenges.

The initial list of items was then sent for review to a group of 8 parents of young children and 11 experts in child development representing the fields of psychology, primary care pediatrics, developmental-behavioral pediatrics, and occupational therapy. Reviewers provided feedback regarding clarity, reading level, and relevance of items. In total, this process resulted in a list of 73 new questions. These questions were screened for Flesch-Kincaid reading level. Items with reading levels over grade 6 were further reviewed and rewritten when possible. The average reading level of the final items was grade 1.8. For each item (e.g. “does your child have a hard time calming down?” or “does your child fight with other children?”), response options were “not at all”, “somewhat” and “very much”.

Study Samples

Participants consisted of parents of children ages 18 months to 5.5 years recruited from primary care practices and referral clinics in the greater Boston area. Primary care sites consisted of 7 urban practices and community health centers and 7 suburban practice groups. Referral sites consisted of 4 developmental-behavioral assessment clinics (including 2 Neonatal Intensive Care Unit (NICU) follow-up programs), 2 child psychiatric clinics, 2 occupational therapy clinics, and 1 speech & language clinic.

We enrolled 3 separate samples:

For scale construction and initial validation,

a) 292 families from pediatric primary care practice groups (hereafter, "*Primary Care Sample*"), and

b) 354 families from referral clinics (hereafter, "*Referral Clinic Sample*").

For replication,

c) 171 families from a separate set of pediatric primary care practices (hereafter, "*Replication Sample*").

Procedures

Parents were enrolled in one of two ways. In settings with high patient volumes (including all primary care and some referral clinics), research assistants approached parents in waiting rooms, described the study, and asked them if they would be interested in participating. In clinics with lower patient volumes (including NICU follow-up and psychiatric clinics), eligible parents were identified from health records. Physicians mailed letters to parents describing the study and stating that a research assistant would call unless the parents indicated their wish not to be contacted (by calling a dedicated voice mail number).

The enrollment process for each procedure is depicted in Figure 1. Among families identified in waiting rooms, 86% enrolled in the study. Complete data were obtained from 71% of enrolled parents (or 61% of potentially eligible parents). Among families identified from medical records, 60% enrolled in the study, and complete data were obtained from 72% of those enrolled (or 43% of those who were eligible).

Assessments

Parents who enrolled were given a packet of questionnaires to complete and mail back. For the *Primary Care* and *Referral Clinic Samples*, the packets consisted of the 73 draft PPSC questions, validated instruments including the CBCL and the ASQ:SE, demographic information, and questions about family risk factors and behavioral diagnoses. The questions about behavioral diagnoses began, "To the best of your knowledge, does your child have…" and included Yes/No responses for "ADHD or ADD," "Anxiety Problem," "Behavior or Conduct Problem," and "Depression." The CBCL is a well-validated, 100-item parent-report checklist that assesses both internalizing behaviors and externalizing behaviors. Scores are provided for both these domains as well as a range of subscales, and clinical cut-offs have been established.[23] The ASQ:SE is an assessment instrument that is commonly used as a screener in pediatric settings. Six separate age-based forms assess children from 18–65 months of age, with length varying from 29 to 36 items.[25] Assessments were similar for the *Replication Sample* with two exceptions: (1) based on initial analyses (see below), an abbreviated form of the PPSC was administered, including 18 final items and 20 exploratory items, and (2) we eliminated the ASQ:SE to reduce participant burden.

Analyses

Four sets of analyses were conducted using Stata version 12[27] and Mplus version 6.11[28]: (1) construction of an abbreviated scale; (2) reliability; (3) screening accuracy; and (4) moderators of scale performance.

1. Construction of an abbreviated scale—To achieve the goal of creating an accurate screening instrument that is feasible for use in primary care, we first calculated descriptive statistics for responses to the 73 questions in the *Primary Care* sample. Then we calculated the frequency of each response category and of missing data, and eliminated items with ≥1% of missing data. We reasoned that more parents from the normative primary care setting would report an absence of each item ("not at all" response) than would report its presence ("somewhat" or "very much" responses); thus we eliminated items that did not follow this pattern.

For the remaining items, we used data from the *Primary Care* and *Referral Clinic* samples to create a bifactor model.[29–31] In a bifactor model, every item loads onto a single general factor, and each item may also load on one or more additional "item clusters" (see Figure 2a).[32] A bifactor model is appropriate because the PPSC is designed to be interpreted based on a single total score. Explicitly modeling a general factor allows us to evaluate the validity of modeling scores along a single dimension, while inclusion of item-clusters helps ensure adequate coverage across domains. To create the initial bifactor model, we conducted exploratory factor analysis (EFA) within a confirmatory framework (CFA) [33]

Once we created an adequate bifactor model of remaining items, we conducted a second stage of analysis to further reduce the number of items. We sought to choose items that yielded the highest loadings on the general factor, that adequately represented the four domains of interest, and that displayed the smallest degree of differential item functioning (DIF) with respect to enrollment site and demographic variables. An item displays differential item functioning if responses differ between two groups after controlling for underlying traits. For example, if children have identical levels of anxiety but parents from two populations (e.g., varying by race, ethnicity, or SES) differ in their likelihood of endorsing an item like "my child is afraid to go outside," then the item can be said to display DIF. We tested for DIF with respect to parent education, ethnicity, race, child gender, child age, and enrollment site.

Once we developed our final bifactor model, we tested its fit to the data in the *Primary Care* and *Referral Clinic* samples, and we also conducted a confirmatory factor analysis in the *Replication Sample*. DIF for each item in the final scale was expressed in terms of an odds ratio (OR). To assess the validity of interpreting a single general factor, we assessed dimensionality by comparing the final bifactor model to two plausible alternative models: (1) a unidimensional model in which all items load on a single factor (similar to the "general factor" in the bifactor model), and (2) a 1st order model with four factors (i.e., a standard factor model as depicted in Figure 2b). We compared each model with regard to factor loadings and adequacy of fit to the data.[32] See Appendix for further detail on methods we used to create an abbreviated scale.

2. Reliability—After selecting a final set of items and constructing a final bifactor model, we calculated Cronbach's alpha for the total PPSC scale in the *Primary Care* and *Referral Clinic* samples, and again for the *Replication Sample*. In addition, 75 participants from the *Replication Sample* were asked to complete the PPSC a second time approximately 4 weeks later. Retest reliability was calculated using an intra-class correlation coefficient (ICC) among the 61 participants (81%) with complete data.

3. Screening accuracy—To establish a clinical threshold, we calculated Receiver Operating Characteristic (ROC) curves in the *Primary Care* and *Referral Clinic Samples*, first using the CBCL as the criterion and then using parent reports of diagnoses as the criterion. After choosing the best overall threshold, we then compared the PPSC's accuracy to the ASQ:SE by calculating each instrument's sensitivity and specificity in regard to (i) CBCL Internalizing, Externalizing, and Total scores; and (ii) parent reports of child diagnoses. Note that sensitivities were calculated for each individual CBCL domain and each individual diagnosis. Because specificity by definition is calculated among individuals who score negative on the criterion, we estimated this parameter only among individuals who scored negative on all CBCL scales, and then among individuals who reported no diagnoses. This prevented individuals who should have scored positive on the PPSC from being included in our calculation of specificity (e.g., if we were to calculate specificity based only on the CBCL's internalizing domain rather than for the CBCL overall, a child with clinically significant externalizing problems would be included in an estimate of specificity, but should not be).

Analyses of the PPSC's sensitivity and specificity with respect to the CBCL and to parent reports of diagnoses were repeated for the *Replication Sample*.

4. Moderators of scale performance—Ideally, the PPSC's total scale does not perform differently across different populations. In addition to assessing DIF while selecting items for the final scale, we tested whether the PPSC's accuracy differs depending on parents' education and children's race, ethnicity, gender, and age. Specifically, we estimated a series of logistic regression models with PPSC scores, a covariate and their interaction as independent variables. Separate sets of analyses were conducted with CBCL score as the dependent variable and parent report of diagnosis as the dependent variable. Significant interaction terms were interpreted as indicating differential accuracy across levels of the covariate.

Results

Characteristics of each of the three samples are reported in Table 1. Across all three samples, the majority of respondents were mothers (82–89%). Most had completed college (52–56%). Family incomes varied, with 31–33% making less than $50,000 per year. The sample was diverse with respect to race and ethnicity, with 25–30% reporting minority race or Hispanic ethnicity. By comparison, 23.9% of Massachusetts residents and 36.3% of U.S. residents reported minority race or Hispanic ethnicity in the 2010 census.[34] Diagnoses of developmental delay and PDD-NOS were reported in 6.6% and 0.7% of the primary care sample, 59.0% and 28.3% of the clinical sample, and 8.5% and 1.1% of the replication sample, respectively.

1. Creation of an abbreviated instrument

We calculated descriptive statistics for each item. Responses for four items were missing for >1% of the sample. For an additional five items, "somewhat" responses were more common than "not at all" responses in the *Primary Care Sample*, suggesting that the items assessed normative behaviors (e.g., "moves around a lot" and "needs a lot of attention"). Items in both categories were dropped.

A 4-factor solution displayed adequate fit in the exploratory factor analysis (EFA), corresponded to our initial hypotheses, and was used to choose initial items for a bifactor model. Specifically, each item cluster in the bifactor model was based on the two items with the highest loadings on each factor. We eliminated 14 items because of low loadings across factors and/or dependencies with other items and built a bifactor model consisting of the 50

remaining items. This model displayed adequate fit to the data (RMSEA = 0.032, CFI = 0.987, TLI = 0.982). The final 18 items were selected by consensus based on their factor loadings, balance among domains of interest, and lack of DIF. This final abbreviated scale was administered to the *Replication Sample*.

The final factor structure displayed adequate fit in the *Primary Care* sample (RMSEA = 0.041, CFI = 0.982, TLI = 0.976), and the *Referral Clinic* sample (RMSEA = 0.045, CFI = 0.987, TLI = 0.984). A confirmatory factor analysis of this model also displayed adequate fit in the *Replication* sample (RMSEA = 0.042, CFI = 0.970, TLI = 0.961). Table 2 presents best estimates of standardized factor loadings and thresholds, calculated on the combined *Primary Care* and *Replication* samples. Note that loading and threshold parameters from a bifactor model with categorical indicators can be directly translated into item-response theory (IRT) difficulty and discrimination parameters.[35]

We assessed DIF for each of the PPSC's final 18 items with respect to parents' education, child's race, ethnicity, age, and gender, yielding a total of 90 separate tests. Among these 90 tests, we identified 11 (12%) that were statistically significant and expressed these in terms of odds ratios (see Table 3). Odds ratios fell above 2 or below 0.5 for five effects (5.6%), indicating relatively large DIF with respect to the given covariates.

We considered whether item clusters provide useful, reliable, and valid information, or whether a single general factor is preferable. We found that 16 of 18 items loaded more strongly on the general factor than on item-clusters. Moreover, the general factor accounted for 50.9% of item variance, while the item clusters accounted for 4.5%, 5.9%, and 3.5%, respectively. As a comparison, we examined a unidimensional model in the same dataset and found that loadings correlated highly with those on the general factor of the bifactor model (r = .93), suggesting that they are closely related. However, the unidimensional model displayed inadequate fit (RMSEA = 0.083, CFI = 0.942, TLI = 0.934), suggesting that item-clusters do enhance the overall model's reliability. In contrast, a 4-factor model based on an EFA also displayed adequate fit (RMSEA=.054, CFI = 0.977, TLI = .972), suggesting that the data may be consistent with either a bifactor model or a multidimensional model. These data support interpretation of the general factor, but highlight the importance of accounting for additional multidimensionality.

2. Reliability

The PPSC total scale was scored by assigning a "0" for each "not at all" response, a "1" for each "somewhat" response, and a "2" for each "very much" response, and then summing the results. Cronbach's alpha for this scale was 0.88 in the *Primary Care* sample, 0.92 in the *Referral Clinic* sample and 0.86 in the *Replication* sample. Retest reliability was adequate (ICC = 0.75).

3. Screening accuracy

Based on ROC analyses of CBCL scores in the *Primary Care* and *Referral Clinic Samples* and of diagnoses in the *Combined Sample*, we found that a cut score of 9 offered the best balance overall between sensitivity and specificity. This cut score was used to calculate sensitivity and specificity with respect to the CBCL and to parents' reports of diagnoses.

a. CBCL—Sensitivity and specificity (and 95% CI) with respect to scores in the clinical range of the CBCL Internalizing domain, the Externalizing domain, Total score, and any one of the three (i.e., "any scale") for *Primary Care* and *Referral Clinic Samples* are presented in Table 4. As described above, specificity was calculated among children who scored positive on none of these scales. Equivalent values for the ASQ:SE are also presented. Sensitivity

and specificity estimates for the PPSC fell consistently above .80. In the *Primary Care Sample*, the PPSC sensitivity was higher than the ASQ:SE's (.88 v .70; $p<.05$), but specificity did not differ. No differences were found between the PPSC and the ASQ:SE in the *Referral Clinic Sample*. The PPSC also displayed sensitivities and specificities over .80 in the *Replication Sample*.

b. Parent reports of diagnoses—Sensitivity and specificity (and 95% C.I.) with respect to parents' reports of child diagnoses of ADHD, anxiety, depression, conduct problems, and any diagnosis are presented in Table 5. Equivalent values for the ASQ:SE and the CBCL are also presented. In the combined *Primary Care* and *Referral Clinic Samples*, sensitivity and specificity estimates for the PPSC consistently fell above .70. No differences were found between the PPSC and the ASQ:SE. In the *Replication Sample*, sensitivity and specificity estimates for the PPSC consistently fell above .70 with the exception of anxiety disorders, for which sensitivity was .33 (.01–.92). The confidence interval for this estimate was extremely large because only 3 children were reported to have anxiety disorders.

4. Moderators of scale performance

We conducted logistic regression analyses to explore whether the accuracy of the PPSC in predicting CBCL scores or parents' reports of diagnoses is moderated by each of the five covariates. Interaction terms were non-significant in all ten analyses, yielding no evidence that the accuracy of PPSC's total scale differs by parent education or by child race, ethnicity, gender, or age.

Discussion

The Preschool Pediatric Symptom Checklist is a new, 18-item instrument designed to screen for social/emotional problems among children 18 months to 5 years of age in primary care settings. It adds to the number of social/emotional screening measures available to primary care pediatricians (e.g., the ASQ:SE, the Brief Infant and Toddler Social Emotional Assessment [BITSEA],[36] the Strengths and Difficulties Questionnaire [SDQ],[37] and the Early Childhood Screening Assessment [ECSA]).[38] The PPSC has the additional benefits of being freely available, brief, easy to read, and easy to score. Because the same questions are posed across the age range, the PPSC is suitable for tracking children's behavior longitudinally. Additionally, it lends itself to simple conversion to electronic formats. Results presented above suggest that despite its brevity, the PPSC's accuracy and reliability are comparable to existing measures.

The sample size of this study was large, diverse in SES and racial/ethnic composition, and was derived from a range of clinical settings. Factor structure was adequate across samples, and a robust general factor supports the use of a total score. Validity was assessed with regard to several criteria and compared directly to another standardized screening instrument. Evidence for differential item functioning was modest, and thus it appears that the PPSC's accuracy does not differ substantially by parents' education or child race, ethnicity, gender or age. The scale shows strong internal and retest reliability, and it identifies children who score in the clinical range of a longer, well-validated and more comprehensive parent-report instrument (the CBCL), as well as children who are reported to have a range of behavioral diagnoses. Moreover, sensitivity and specificity with respect to these criteria were comparable to another well-accepted screener, the ASQ:SE. Finally, results for the PPSC's total scale remained consistent when replicated in an independent sample.

We recommend use of the total score, using a cutoff score of 9, and caution against interpreting item clusters. We accounted for several dimensions to ensure adequate coverage

of internalizing, externalizing, attention symptoms, and parenting challenges, and to improve model fit. However, given the small number of items per cluster, they are likely to be unreliable, and the relative dominance of the general factor makes it likely that the total score is a better predictor of each type of disorder than a subscale score would be.[39]

We note several limitations to our study. First, our sample consisted of a convenience sample of parents who brought their children to pediatricians' offices or to developmental specialists in the greater Boston area, and thus the results of this study generalize most readily to the practices where the data were collected. The 14 primary care practices and 9 referral clinics from which we recruited participants were highly varied in their practice models and in the SES and ethnic backgrounds of their clientele. To maximize the representativeness of our sample, we approached all parents of young children, and used a random-number system to choose families when waiting rooms were crowded.

We were not able to enroll all parents who sought pediatric care for their children. From waiting rooms, 61% of eligible parents both enrolled and completed study materials. Among families identified through medical records, 159 families could not be contacted (30% of the sample). We may have failed to reach these parents because they did not want to participate, or because their contact information had changed—it is impossible to assess which. Thus, only 42% of families identified through medical records enrolled and completed study materials. This non-response bias may have affected results in that parents of children with severe behavioral difficulties may have been less likely to participate due to their increased stress and pressures on their time and patience. Alternatively, parents whose children showed the least behavioral difficulties may also have been less likely to enroll as they may not have appreciated the need for behavioral screening. Despite the relatively high burden on participants imposed by our study and the lack of clinical relevance to each participant (individual results were not reported to families or pediatricians), our enrollment rates were comparable to screening rates reported in other studies.[40, 41]

We utilized parents' CBCL scores and their reports of existing diagnoses as our criterion measures, rather than structured diagnostic interviews. Ideally, such interviews would have been conducted by clinicians who were unaware of each participant's screening results. Nevertheless, even structured diagnostic interviews rely on parent proxies for children in this age range, and therefore the results of all social/emotional assessment measures for young children should be interpreted with caution. Nonetheless, early identification and treatment offer significant benefits for young children with social/emotional disorders and the PPSC appears to be an accurate and feasible tool that can be used to improve access to services.

Summary

We have presented data regarding the development and validation of the Preschool Pediatric Symptom Checklist (PPSC), a new screening instrument for use in pediatric primary care settings to assess preschool children's social/emotional development. The PPSC is one part of a comprehensive pediatric surveillance instrument for children less than 5 years of age: the Survey of Wellbeing of Young Children (SWYC). Development and validation of other segments of the larger instrument are ongoing and will be reported in subsequent publications.

Acknowledgments

We are indebted to The Commonwealth Fund for its support of the larger project of which this is a part. In particular, Dr. Ed Schor's vision and imagination motivated us to create a practical instrument to guide surveillance in pediatric primary care settings. Dr. Michael Jellinek had the foresight long ago to create a tool to help pediatricians identify children who had unmet needs for mental health attention. We are grateful also for the

assistance of Dr. Robin Adair, Dr. Liz McGowan, and a large number of pediatricians and other child health specialists[1], who allowed us to enroll parents through their offices, to Kate Mattern and Robyn Della Giustina for help with data collection and management, and to the Tufts Medical Center CTSI for ongoing support. We acknowledge as well the wise suggestions of anonymous reviewers of an earlier version of this paper.

Research support for the development and validation of the PPSC was provided by The Commonwealth Fund & NIH Grant KM1CA156726

References

1. Egger HL, Angold A. Common emotional and behavioral disorders in preschool children: presentation, nosology, and epidemiology. Journal of Child Psychology & Psychiatry. 2006; 47(3–4):313–337. [PubMed: 16492262]
2. Brauner CB, Stephens CB. Estimating the prevalence of early childhood serious emotional/behavioral disorder: challenges and recommendations. Public Health Rep. 2006; 121(3):303–310. [PubMed: 16640155]
3. Briggs-Gowan MJ, Carter AS. Social-emotional screening status in early childhood predicts elementary school outcomes. Pediatrics. 2008; 121(5):957–962. [PubMed: 18450899]
4. O'Connell, ME.; Boat, T.; Warner, KE., editors. Preventing Mental, Emotional, and Behavioral Disorders Among Young People: Progress and Possibilities. Washington, DC: National Academies Press; 2009. Youth and Young Adults: Research Advances and Promising Interventions Committee on the Prevention of Mental Disorders and Substance Abuse Among Children, Institute of Medicine, National Research Council.
5. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age of onset distributions of DSM-IV disorders in the National Comorbidity Study Replication. Archives of General Psychiatry. 2005; 62(6):593–602. [PubMed: 15939837]
6. Sheldrick RC, Merchant S, Perrin EC. Identification of developmental-behavioral problems in primary care; a systematic review. Pediatrics. 2011; 128(2):356–363. [PubMed: 21727101]
7. Sheldrick RC, Perrin EC. Surveillance of Children's Behavior and Development: Practical Solutions for Primary Care. Journal of Developmental and Behavioral Pediatrics. 2009; 30(2):151–153. [PubMed: 19363367]
8. Pediatric Symptom Checklist [PSC]. Massachusetts General Hospital Website, Child Psychiatry. 2012. Retrieved from: http://www2.massgeneral.org/allpsych/psc/psc_home.html
9. Jellinek M, Murphy JM, Burns B. Brief psychosocial screening in outpatient pediatric practice. Pediatrics. 1986; 109:371–378.
10. Simonian SJ, Tarnowski KJ. Utility of the Pediatric Symptom Checklist for behavioral screening of disadvantaged children. Child Psychiatry and Human Development. 2001; 31(4):269–278. [PubMed: 11227987]
11. Jutte DP, Burgos A, Mendoza F, Ford CB, Huffman LC. Use of the Pediatric Symptom Checklist in a low-income, Mexican American population. Archives of Pediatrics and Adolescent Medicine. 2003; 157(12):1169–1176. [PubMed: 14662568]
12. Gardner W, Lucas A, Kolko DJ, Campo JV. Comparison of the PSC-17 and alternative mental health screens in an at-risk primary care sample. Journal of the American Academy of Child & Adolescent Psychiatry. 2007; 46(5):611–618. [PubMed: 17450052]
13. Jellinek MS, Murphy JM, Little M, et al. Use of the Pediatric Symptom Checklist (PSC) to screen for psychosocial problems in pediatric primary care: A national feasibility study. Archives of Pediatrics and Adolescent Medicine. 1999; 153(3):254–260. [PubMed: 10086402]

[1]Primary Care Clinics: Harvard Vanguard Medical Associates, Medical Associates Pediatrics, Revere Health Care Center, Pediatric Healthcare Associates, Porter Pediatrics, Southborough Medical Group, Floating Hospital General Pediatrics, Wilmington Pediatrics, Quality Kids Care, Codman Square Health Center, Westwood- Mansfield Pediatric Associates, Dr. Babu Pediatrics.
Referral Clinics: Floating Hospital Specialty Clinics, including the Center for Children with Special Needs, NICU Follow-Up clinic, Child Psychiatry, and International Adoption Clinic; Occupational Therapy Associates Watertown, South Shore Therapies; Massachusetts General Hospital Child Psychiatry; University of Massachusetts NICU Follow-Up & Developmental-Behavioral Pediatrics Clinics; Boston Medical Center Developmental Assessment Clinic, Harvard Vanguard Speech/Language Pathology.

14. Jellinek, MS.; Froehle, MD. Bright futures in practice: Mental health. Arlington, VA: National Center for Education in Maternal and Child Health; 1998.
15. American Academy of Pediatrics Task Force on Mental Health. Addressing mental health concerns in primary care: A clinician's toolkit. Pediatrics. 2010; 125(Supp 3):S173–S192.
16. National Quality Forum. NQF Endorses Child Health Outcome Measures. 2011 [Press Release]. Retrieved from http://www.qualityforum.org/News_And_Resources/Press_Releases/2011/NQF_Endorses_Child_Health_Outcome_Measures.aspx;
17. Murphy, JM.; Jellinek, MS. Rockville, MD: 1985. Development of a brief psychosocial screening instrument for pediatric practice: final report. NIMH contract No. 84M0313612
18. Jellinek MS, Murphy JM, et al. The pediatric symptom checklist: screening school-age children for psychological dysfunction. Pediatrics. 1988; 112:201–209.
19. Murphy JM, Jellinek MS. Screening for psychosocial dysfunction in economically disadvantaged and minority group children: further validation of the Pediatric Symptom Checklist. Amer J of Orthopsychiatry. 1988; 58(3):450–456. [PubMed: 3407735]
20. Little M, Murphy JM, Jellinek MS, Bishop SJ, Arnett HL. Screening four and five year old children for psychosocial dysfunction; A preliminary study with the Pediatric Symptom Checklist. Journal of Developmental and Behavioral. Pediatrics. 1994; 15:191–197.
21. Pagano M, Murphy JM, Pederson M, et al. Screening for psychosocial problems in 4–5 year olds during routine EPSDT examinations; Validity and reliability in Mexican-American sample. Clinical Pediatrics. 1996; 35:139–146. [PubMed: 8904487]
22. Kelleher KJ, Childs GE, Wasserman RC, McInerny TK, Nutting PA, Gardner WP. Insurance status and recognition of psychosocial problems. A report from the Pediatric Research in Office Settings and the Ambulatory Sentinel Practice Networks. Archives of Pediatrics & Adolescent Medicine. 1998; 151(11):1109–1115. [PubMed: 9369872]
23. Achenbach, TM. Manual for the Child Behavior Checklist/4–18 and 1991 Profile. Burlington, VT: Department of Psychiatry, University of Vermont; 1991.
24. Briggs-Gowan MJ, Carter AS, Irwin JR, Wachtel K, Cicchetti DV. The brief infant-toddler social and emotional assessment: Screening for social-emotional problems and delays in competence. Journal of Pediatric Psychology. 2004; 29:143–155. [PubMed: 15096535]
25. Squires, J.; Bricker, D.; Twombley, E. The ASQ:SE User's Guide. Baltimore, MD: Paul H Brookes Publishing Co.; 2002.
26. Stoppelbein L, Greening L, Jordan SS, Elkin TD, Moll G, Pullen J. Factor analysis of the Pediatric Symptom Checklist with a chronically ill pediatric population. Journal of Developmental and Behavioral Pediatrics. 2005; 26:349–355. [PubMed: 16222174]
27. StataCorp. Stata Statistical Software: Release 12. College Station, TX: StataCorp LP; 2011.
28. Muthén, LK.; Muthén, BO. Mplus Version 6.11. Los Angeles, CA: Muthén & Muthén; 2011.
29. Patrick CJ, Hicks BM, Nichol PE, Krueger RF. A bifactor approach to modeling the structure of the Psychopathy Checklist-Revised. Journal of Personality Disorders. 2007; 21:118–141. [PubMed: 17492917]
30. Kruger RF, Markon KE, Patrick CJ, Benning SD, Kramer MD. Linking antisocial behavior, substance abuse, and personality: An integrative quantitative model of the adult externalizing spectrum. Journal of Abnormal Psychology. 2007; 116:645–666. [PubMed: 18020714]
31. Gibbons RD, Rush AJ, Immekus JC. On the psychometric validity of the domains of the PDSQ: An illustration of the bi-factor item response theory model. Journal of Psychiatric Research. 2009; 43:401–410. [PubMed: 18554611]
32. Reise SP, Morizot J, Hays RD. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. Quality of Life Research. 2007; 16:s19–s31.
33. Muthen LK, Muthen B. Mplus short course: Topic 1: Exploratory factor analysis, confirmatory factor analysis, and structural equation modeling for continuous outcomes. 2009 Retrieved from: http://www.statmodel.com/download/Topic%201.pdf.
34. [accessed Feb 7, 2012] U.S. census bureau webpage. http://quickfacts.census.gov
35. Gibbons RD, Bock RD, Hedeker D, Weiss DJ, Segawa E, Bhaumik DK, et al. Full-information item bifactor analysis of graded response data. Applied Psychological Measurement. 2007; 31:4–19.

36. Briggs-Gowan, MJ.; Carter, AS. The Brief Infant-Toddler Social and Emotional Assesment (BITSEA) manual. New Haven, CT: Department of Psychology, Yale University; 2001.

37. Goodman R. Psychometric properties of the Strength and Difficulties Questionnaire. Journal of the American Academy of Child and Adolescent Psychiatry. 2001; 40:1337–1345. [PubMed: 11699809]

38. Gleason MM, Zeanah CH, Dickstein S. Recognizing young children in need of mental health assessment: Development and preliminary validity of the early childhood screening assessment. Infant Mental Health Journal. 2010; 31(3):335–357.

39. Reise SP, Moore TM, Haviland MG. Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. Journal of Personality Assessment. 2010; 92:544–559. [PubMed: 20954056]

40. Hix-Small H, Marks K, Squires J, Nickel R. Impact of implementing developmental screening at 12 and 24 months in a pediatric practice. Pediatrics. 2007; 120(2):381–389. [PubMed: 17671065]

41. Schonwald A, Huntington N, Chan E, Risko W, Bridgemohan C. Routine developmental screening implemented in urban primary care settings: More evidence of feasibility and effectiveness. Pediatrics. 2009; 123(2):660–668. [PubMed: 19171635]

42. Hu L, Bentler P. Fit indices in covariance structure analysis: Sensitivity to underparameterized model misspecifications. Psychological Methods. 1998; 4:424–453.

43. Kline, RB. Principles and practice of structural equation modeling. New York: Guilford Press; 1998.

44. Teresi JA, Fleishman JA. Differential item functioning and health assessment. Quality of Life Research. 2007; 16:33–42. [PubMed: 17443420]

45. Yang FM, Tommet D, Jones RN. Disparities in self-reported geriatric depressive symptoms due to sociodemographic differences: An extension of the bi-factor item response theory model for use in differential item functioning. Journal of Psychiatric Research. 2009; 43:1025–1035. [PubMed: 19211113]

46. Cole SR, Kawachi I, Maller SJ, Berkman LF. Test of item-response bias in the CESD scale. Experience from the New Haven EPESE Study. Journal of Clinical Epidemiology. 2000; 53:285–289. [PubMed: 10760639]

Appendix. Details regarding analyses to construct an abbreviated scale

We conducted latent variable modeling in Mplus with the goal of creating a bifactor model of a final scale. To create the bifactor model, we used Mplus software to conduct exploratory factor analysis (EFA) in a confirmatory framework.[33] Specifically, we conducted an EFA to identify factors and used a Promax rotation to identify items that loaded most strongly on each factor. This information was used to create an initial confirmatory factor analytic (CFA) model in which all items loaded on a general factor and each item-cluster was defined by the two items with the strongest loadings in the EFA. To achieve model identification, loadings of these items were constrained to zero on other factors and factor means and variances were constrained to 0 and 1, respectively. Covariances among factors were also constrained to 0. To further refine the model, we relied on Mplus “modification indices” which estimate the impact of adding individual parameters to the model. Using an iterative process, we added items to factors one at a time, choosing those items that offered the greatest improvement in model fit and were consistent with prior theory (i.e., question content was consistent with the item cluster). Each time a parameter was added to the model, model fit was re-estimated. We continued adding parameters until we achieved a bifactor model of as many questions as possible that displayed adequate fit. At this stage, items were eliminated if they displayed very low loadings on both the general factor and on item-cluster factors, or if they were highly similar to another item that displayed superior fit.

Unless otherwise specified, all models were estimated using robust weighted least squares (WLSMV) to account for any violations of multivariate normality in the data. Note that

loading and threshold parameters from a bifactor model with categorical indicators can be directly translated into item-response theory (IRT) difficulty and discrimination parameters.[32]

Model fit was evaluated using standard indices. Specifically, we strove to develop models with Root Mean Square Errors of Approximation (RMSEA) <.06, Comparative Fit Index (CFI) > .95, and Tucker Lewis Index (TLI) > .95.[42] Because adding parameters will almost always improve model fit, each of these indices includes a penalty for increased complexity, thus favoring parsimonious models. Although we considered interpreting non-significant tests of model chi-square as an indicator of model fit, we excluded this index because such tests are considered over-powered in large samples.[43]

Once we had an adequate bifactor model of remaining items, we conducted a second stage of analysis to further reduce items. In this stage, we sought to choose items that displayed the highest loadings on the general factor, that adequately represented the four domains of interest, and that displayed the smallest degree of differential item functioning (DIF) with respect to enrollment site and demographic variables. An item displays differential item functioning if responses differ between two groups only after controlling for underlying traits. For example, if children have identical levels of anxiety but parents from two populations differ in their likelihood of endorsing an item like “my child is afraid to go outside,” then the item can be said to display DIF. To identify items with strong DIF, we added several covariates to create a model commonly known as a Multiple Indicator Multiple Causes (MIMIC) model, where indicators are questions and “causes” are covariates. Although MIMIC models only detect uniform and not non-uniform DIF (i.e., they only detect DIF to the extent that its magnitude is proportional across levels of the underlying trait), we chose a MIMIC model because of its power to detect DIF across multiple covariates simultaneously.[44] In our model, all latent variables were regressed on six covariates. The six binary covariates were (1) parent education (high school education or less versus education beyond high school), (2) ethnicity (Hispanic versus not), (3) race (white versus non-white), (4) child gender (male versus female), (5) child age (18 months – 3 years versus >3 years), and (6) enrollment site (primary care versus referral clinic). To control for underlying traits, all latent variables were regressed on the six covariates, and these covariates were allowed to covary. Modification indices were then used to identify items on which one or more of the six covariates showed strong direct effects.

After we had identified a final subset of items, we repeated the process of model building to ensure optimal factor structure. To ensure that the final model would display adequate fit in both the *Primary Care* and *Referral Clinic* samples, this analysis was conducted with each sample identified as a separate group using the Mplus “mgroup” function and all factor loadings and thresholds constrained across groups. To evaluate factor structure, we tested the final bifactor model across *Primary Care* and *Referral Clinic* samples using the fit indices described above. We also conducted a confirmatory factor analysis in the *Replication* sample.

Although items with relatively large DIF were eliminated during the process of creating an abbreviated scale, some degree of DIF could remain in the final set of items. To characterize the degree of DIF in our final model, we used the combined *Primary Care* and *Replication* samples to test for uniform DIF using a latent variable MIMIC model. Following recommendations of Yang et al (2009),[45] we used Mplus modification indices to identify the item on which a covariate had the largest direct effect, controlling for latent variable scores. We then included a direct effect between this item and the covariate, re-estimated the model, and conducted a difference test to determine whether adding the parameter made a significant improvement in model fit. We continued to add direct relationships between

items and covariates in this stepwise fashion until the difference test between successive models was non-significant. Next, we re-estimated the final model using a robust maximum likelihood estimator with a logit link, which allows direct effects to be expressed in terms of odds ratios (OR). Following recommendations of Cole et al (2000),[46] we used a threshold of OR > 2.0 or OR < 0.5 to classify DIF as "relatively large."

To evaluate the appropriateness of interpreting the general factor as a primary dimension, we followed a set of procedures recommended by Reise et al.[32] For example, we compared factor loadings and variance explained in the bifactor model between the general factor and the item-cluster factors.[32] Stronger loadings on the general factor and greater variance explained by the general factor support its interpretation as a primary dimension. We also compared the final bifactor model to two alternative model specifications: a unidimensional model and a first-order 4-factor model created based on a separate EFA. Comparisons were made in regard to model fit indices and factor loadings.

In summary, we used the methods described above to (1) select items for the PPSC, and (2) evaluate the final factor model with respect to fit, DIF, and dimensionality.

What's New

The PPSC is a brief social/emotional screening instrument designed for use in pediatrics. It is easy to administer and score, and freely available. Initial investigation suggests that it has sound psychometric properties and effectively identifies children with social/emotional problems.

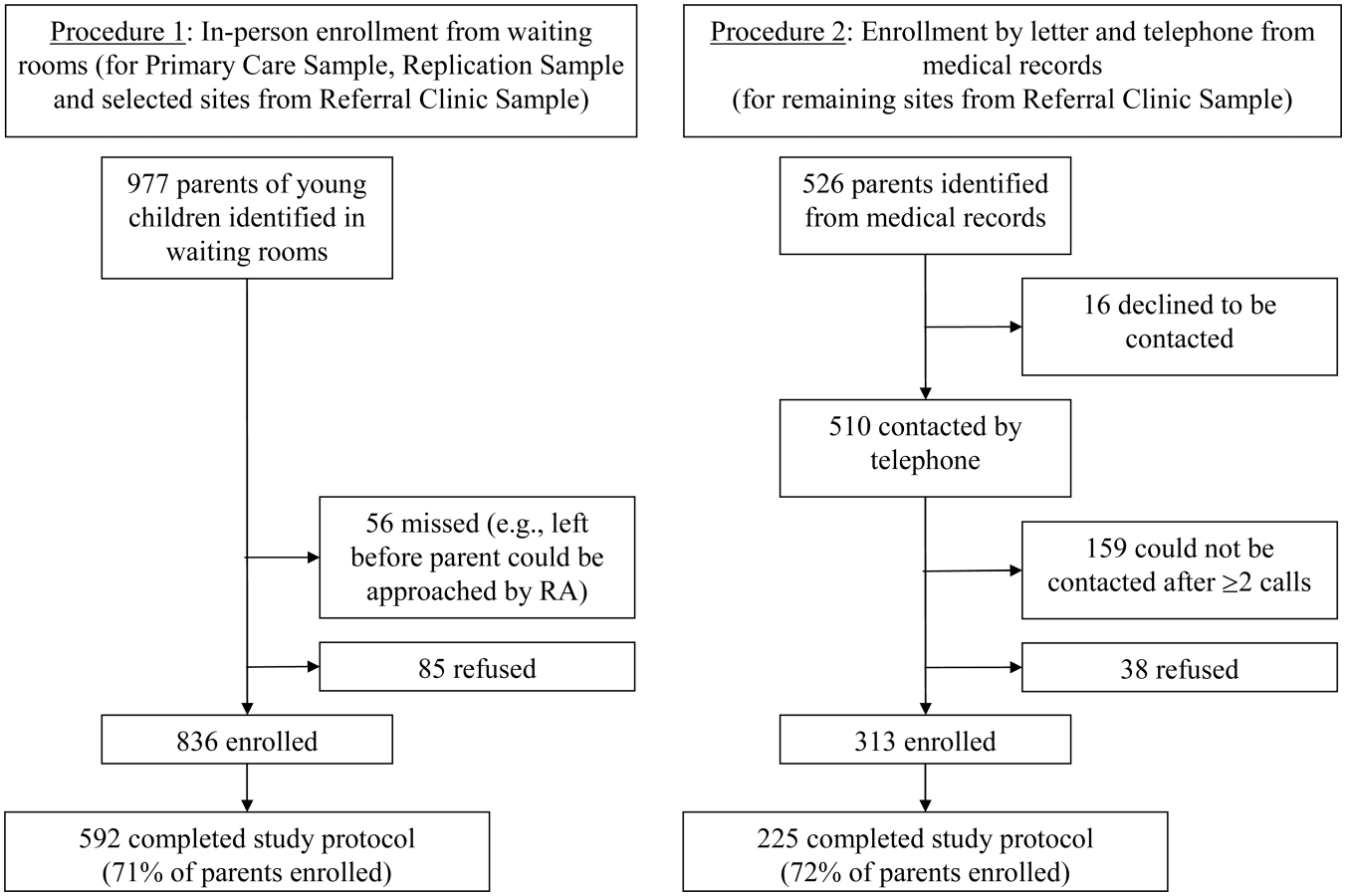

Figure 1.
Enrollment

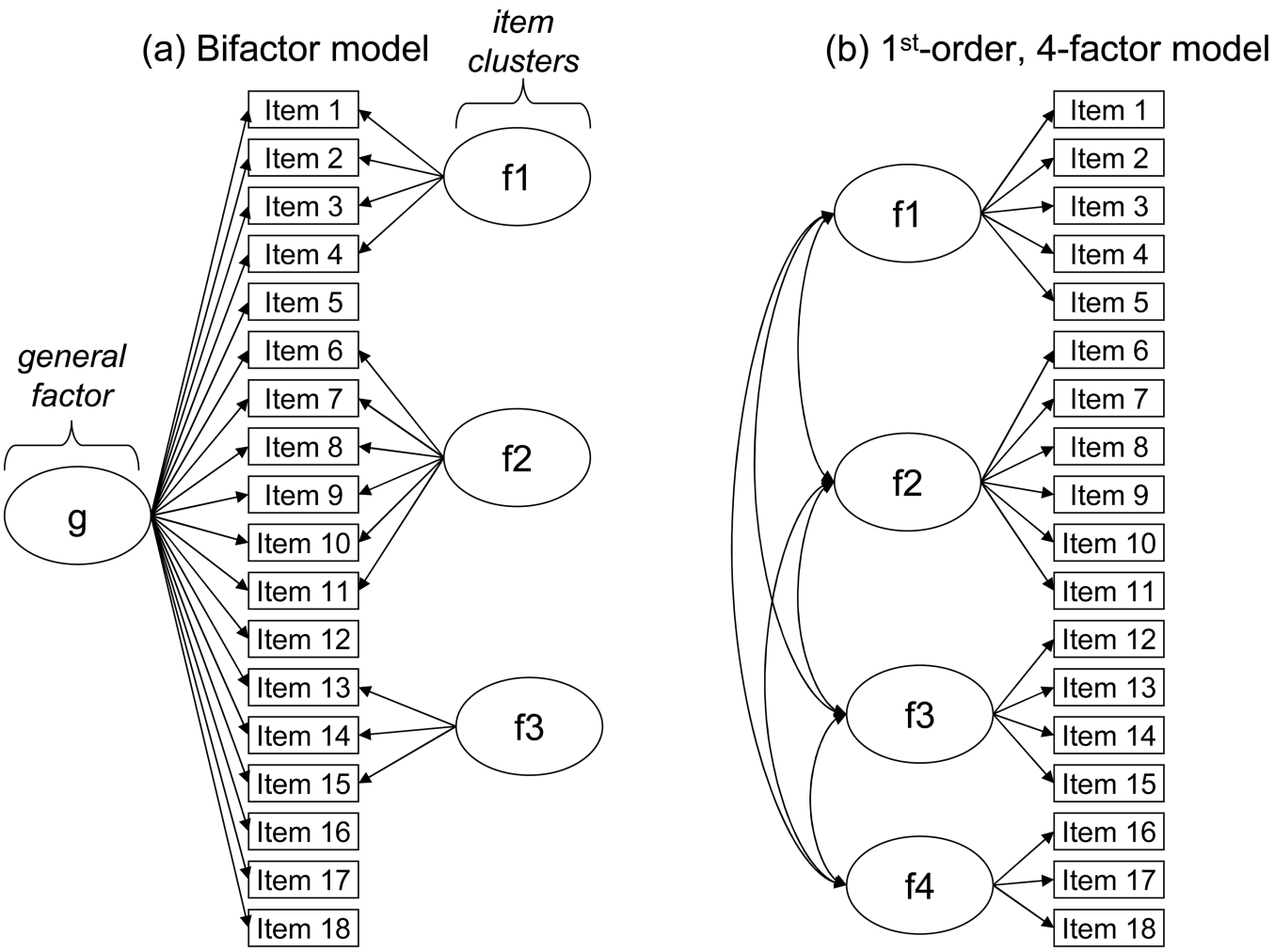

Figure 2.
Examples of alternative model specifications

Table 1

Sample characteristics

	Primary Care Sample		Referral Clinic Sample		Replication Sample	
n	292		354		171	
Child Male	161	55%	221	62%	97	57%
Child Hispanic	39	13%	47	13%	21	12%
Child Race						
Caucasian	214	73%	267	75%	126	74%
African-American	35	12%	40	11%	29	17%
Native American	3	1%	10	3%	0	0%
Asian	20	7%	19	5%	9	5%
Other/multiple races	9	3%	10	3%	5	3%
not indicated	11	4%	8	2%	2	1%
Child age in months						
18–23	50	17%	40	11%	30	18%
24–29	58	20%	46	13%	28	16%
30–35	31	11%	46	13%	26	15%
36–47	83	28%	106	30%	46	27%
48–59	42	14%	78	22%	28	16%
60+	28	10%	38	11%	12	7%
Premature birth						
(<37 wks)	45	15%	162	46%	16	9%
Public health insurance	59	20%	94	27%	32	19%
Mother completed forms	249	85%	315	89%	141	82%
Parent Education						
<High School	13	4%	12	3%	7	4%
High School diploma	75	26%	79	22%	40	23%
some college	47	16%	62	18%	30	18%
College diploma	100	34%	111	31%	45	26%
advanced degree	57	20%	87	25%	45	26%

	Primary Care Sample		Referral Clinic Sample		Replication Sample	
Family Income						
< $20,000	40	14%	60	17%	35	20%
$20,000–49,999	51	17%	51	14%	23	13%
$50,000–99,999	81	28%	117	33%	36	21%
> $100,000	112	38%	113	32%	35	20%
not indicated	8	3%	13	4%	42	25%
Clinic Type						
Primary Care	292	100%	--	--	171	100%
Devel-Behav Ped	--	--	156	44%	--	--
Psychiatry	--	--	21	6%	--	--
NICU follow-up	--	--	119	34%	--	--
Occupational therapy	--	--	50	14%	--	--
Speech & Language	--	--	8	2%	--	--

Table 2

Final bifactor model

Item	Standardized factor loadings*				Thresholds**	
	General factor	Externalizing item-cluster	Internalizing item-cluster	Attention problems cluster	between "not at all" and "somewhat"	between "somewhat" and "very much"
1. break things on purpose?	0.670	0.342			0.928	1.892
2. fight with other children?	0.560	0.627			0.773	2.063
3. aggressive?	0.716	0.409			0.635	1.706
4. angry?	0.750	0.364	0.223		1.061	1.991
5. have trouble playing with other children?	0.746		0.254		0.511	1.622
6. seem sad or unhappy?	0.639		0.273		1.294	2.506
7. seem nervous or afraid?	0.533		0.489		0.823	2.090
8. get upset if things are not done in a certain way?	0.624		0.468		0.009	1.245
9. have a hard time with change?	0.714		0.498		0.336	1.299
10. have trouble paying attention?	0.818			0.418	0.020	1.184
11. fidgety or unable to sit still?	0.742			0.373	0.142	1.060
12. have trouble staying with one activity?	0.699			0.581	0.107	1.118
13. take your child out in public?	0.855				0.653	1.587
14. get your child to obey you?	0.844				0.000	1.409
15. comfort your child?***	0.761				1.072	2.014
16. know what your child needs?***	0.644				0.754	1.619
17. keep your child on a schedule or routine?***	0.650				0.700	1.609
18. have a hard time calming down?	0.847				0.058	1.228

Note. Model estimated in the combined primary care and replication samples; To account for the PPSC's 3-option (non-continuous) response scale, we modeled responses to all PPSC items as categorical variables.

* With categorical indicators, factor loadings can be interpreted as in standard factor models.

** With categorical indicators, thresholds are estimated for each item rather than intercepts. The first threshold indicates the score on the latent variable at which the most likely observed response shifts from 0 ("not at all") to 1 ("somewhat"), while the second threshold indicates the latent score at which the most likely observed response shifts from 1 (i.e., "somewhat") to 2 (i.e., "very much").

*** Because each factor must be modeled by a minimum of three items and loadings on the "parenting challenges" factor were comparatively modest, we chose to recast dependencies among these items by including covariances among individual items. Because the three items that defined "attention problems", displayed higher loadings, we maintained this item-cluster as a factor in our model. Item 16 covariance with item 15 = 0.33 and with item 17 = 0.22

Table 3

Differential item functioning (DIF) as evidenced by odds ratios of direct effects of covariates on items

	Covariate			
	white race	**Hispanic ethnicity**	**education: high school or less**	**age: 3.2 years or less**
1. break things on purpose?			2.18*	
4. angry?		2.36*		
5. have trouble playing with other children?	1.36			
6. seem sad or unhappy?			0.68	
10. hard to comfort your child?	0.77			
11. have trouble paying attention?	1.97			
13. have trouble staying with one activity?				2.3*
16. hard to know what your child needs?*	0.62			
17. hard to keep your child on a schedule or routine?	0.45*			
18. have a hard time calming down?	1.51			0.41*

Note.

* indicates effects considered "relatively large"; No effects were found for covariate male gender; no evidence of DIF was found for items not listed

Table 4

Sensitivity and Specificity (with 95% C.I.) with respect to CBCL 1.5–5

		Primary Care Sample					Referral Clinic Sample					Replication Sample		
			PPSC		ASQ-SE			PPSC		ASQ-SE			PPSC	
		f(pos)	Sensitivity	Specificity	Sensitivity	Specificity	f(pos)	Sensitivity	Specificity	Sensitivity	Specificity	f(pos)	Sensitivity	Specificity
CBCL 1.5–5 clinical range	CBCL Total Score	16	0.93 (0.66,1.0)		0.87 (0.6,0.98)		87	0.99 (0.94,1)		0.98 (0.92,1)		9	1 (0.66,1.0)	
	CBCL Internalizing Domain	19	0.88 (0.64,0.99)		0.72 (0.47,0.9)		94	0.93 (0.86,0.98)		0.88 (0.79,0.94)		10	0.90 (0.55,1.0)	
	CBCL Externalizing Domain	19	0.94 (0.71,1.0)		0.79 (0.54,0.94)		80	0.99 (0.93,1.0)		0.96 (0.89,0.99)		9	0.88 (0.47,1.0)	
	Any CBCL Domain[†]	28	0.88 (0.7,0.98)	0.89 (0.85,0.93)	0.70 (0.5,0.86)	0.93 (0.89,0.96)	116	0.94 (0.88,0.97)	0.77 (0.71,0.82)	0.88 (0.81,0.94)	0.71 (0.64,0.77)	10	0.87 (0.6,0.98)	0.89 (0.83,0.94)

Note. f(pos) = frequency of positive results on CBCL.

[†]"Any Domain" refers to positive scores on the Internalizing Domain, the Externalizing Domain, and/or the Total Score. Because children who scored positive on any domain of the CBCL were also expected to score positive on the PPSC, specificity is calculated only for this category

Table 5

Sensitivity and Specificity (with 95% C.I.) with respect to parent report of behavioral diagnoses

		Combined Sample Primary Care + Referral Clinics)					Replication Sample		
		f(diag)	PPSC		ASQ-SE		f(diag)	PPSC	
			Sensitivity	Specificity	Sensitivity	Specificity		Sensitivity	Specificity
Parent report of:	ADHD	36	0.89 (0.73,0.97)		0.83 (0.66,0.93)		5	1.00 (0.15,1)	
	Anxiety	44	0.76 (0.60,0.88)		0.72 (0.56,0.85)		3	0.33 (0.01,0.91)	
	Behavior or Conduct Problems	80	0.92 (0.84,0.97)		0.87 (0.77,0.94)		16	1.00 (0.59,1)	
	Depression	4	0.75 (0.19,0.99)		0.50 (0.07,0.93)		0	N/A	
	Any Diagnosis	117	0.86 (0.78,0.92)	0.77 (0.73,0.81)	0.80 (0.71,0.87)	0.82 (0.79,0.85)	24	0.82 (0.48,0.97)	0.87 (0.81,0.92)

Note. f(diag) = frequency of each diagnosis in each sample

†"Any Diagnosis" refers to diagnoses of ADHD, Anxiety, Behavior or Conduct Problems and/or Depression. Because children with any one of these diagnoses were also expected to score positive on the PPSC, specificity is calculated only for this category.