Developmentally and Culturally Appropriate Screening in Primary Care: Development of the Behavioral Health Checklist

Thomas J. Power,[1,2] PhD, Anson J. Koshy,[1] MD, Marley W. Watkins,[3] PhD, Michael C. Cassano,[1,4] PhD, Andrea C. Wahlberg,[1,5] MEd, Jennifer A. Mautone,[1] PhD, and Nathan J. Blum,[1,2] MD

[1]*Department of Pediatrics, and Department of Child and Adolescent Psychiatry and Behavioral Sciences, The Children's Hospital of Philadelphia,* [2]*Departments of Pediatrics and Psychiatry, Perelman School of Medicine at University of Pennsylvania,* [3]*Department of Educational Psychology, Baylor University* [4]*Institute for Graduate Clinical Psychology, Widener University, and* [5]*Department of Educational Psychology, Texas A&M University*

All correspondence concerning this article should be addressed to Thomas J. Power, PhD, CHOP North (3535 Market) - Room 1471, The Children's Hospital of Philadelphia, 34th St. and Civic Center Blvd., Philadelphia, PA 19104, USA. E-mail: power@email.chop.edu

Received April 8, 2013; revisions received July 18, 2013; accepted July 21, 2013

Objective To evaluate the construct validity of the Behavioral Health Checklist (BHCL) for children aged from 4 to 12 years from diverse backgrounds. **Method** The parents of 4–12-year-old children completed the BHCL in urban and suburban primary care practices affiliated with a tertiary-care children's hospital. Across practices, 1,702 were eligible and 1,406 (82.6%) provided consent. Children of participating parents were primarily non-Hispanic black/African American and white/Caucasian from low- to middle-income groups. Confirmatory factor analyses examined model fit for the total sample and subsamples defined by demographic characteristics. **Results** The findings supported the hypothesized 3-factor structure: Internalizing Problems, Externalizing Problems, and Inattention/Hyperactivity. The model demonstrated adequate to good fit across age-groups, gender, races, income groups, and suburban versus urban practices. **Conclusion** The findings provide strong evidence of the construct validity, developmental appropriateness, and cultural sensitivity of the BHCL when used for screening in primary care.

Key words behavior problems; mental health; primary care.

Mental health problems are among the most prevalent chronic conditions of youth (Van Cleave, Gortmaker, & Perrin, 2010). The American Academy of Pediatrics has asserted that primary care providers have an important role in providing services to children with mental health concerns (Foy, Kelleher, Laraque, & American Academy of Pediatrics Task Force on Mental Health., 2010). As a first step, it is critical for primary care providers to screen for behavioral and emotional concerns to identify children who may need further evaluation and intervention services (Bell, Johnson, Myers, & Patrick, 2010; Navon, Nelson, Pagano, & Murphy, 2001). The American Academic of Pediatrics Task Force on Mental Health (2010) recommends the screening of children and adolescents (beginning at 5 years) for behavioral and psychosocial problems at routine health maintenance visits as well as when psychosocial concerns arise.

Major disparities have been identified in the provision of health and mental health services to children across the broad range of ethnic/racial groups and socioeconomic levels represented in the United States (Braveman & Barclay, 2009). For example, researchers have documented significant ethnic and racial disparities in service delivery for children with attention-deficit/hyperactivity disorder (ADHD) and other mental health conditions; children of minority status have been substantially

Journal of Pediatric Psychology 38(10) pp. 1155–1164, 2013
doi:10.1093/jpepsy/jst063
Advance Access publication August 26, 2013

underdiagnosed and undertreated (Flores & the Committee on Pediatric Research, 2010). Reducing health disparities requires comprehensive systemic efforts at multiple levels of the service delivery system (Braveman & Barclay, 2009), but one strategy to address this problem is to develop screening and assessment tools that are valid for use with children from diverse racial, ethnic, and socioeconomic populations.

Parent-report instruments are commonly used to screen children for behavioral health concerns; yet, existing instruments are limited with regard to their inclusion of factors that are theoretically and clinically meaningful, their validity with diverse populations, and their feasibility for use in primary care. The Pediatric Symptom Checklist-17 (PSC-17) and the Strengths and Difficulties Questionnaire (SDQ) are the most well-researched and commonly used multidimensional scales for behavioral health screening in primary care.

The PSC-17 was developed as a parent-report screening measure of externalizing, internalizing, and attention problems for children aged 4–17 years (Gardner et al., 1999). Studies to develop and evaluate the psychometric properties of this instrument have been conducted primarily with a non-Hispanic white population (Gardner, Lucas, Kolko, & Campo, 2007; Gardner, et al., 1999). A recent study provided evidence to support the validity of the PSC-17 with both white/Caucasian and black/African American samples (Stoppelbein, Greening, Moll, Jordan, & Suozzi, 2012). However, another study using a relatively large sample of predominantly black/African American low-income children ($n = 320$) served through urban primary care practices raised questions about the construct validity of the PSC-17 (Kostanecka et al., 2008). Specifically, many of the items, particularly on the externalizing and attention problems subscales, had significant loadings on more than one factor, and the three-factor model failed to meet statistical criteria for adequate model fit. In addition, the internalizing scale of the PSC-17 is limited in screening for anxiety, given that there is only one item that specifically refers to anxiety. Further, the PSC-17 uses one set of items across the preschool and adolescent years, even though there is considerable evidence of developmental variations in internalizing, externalizing, and attention problems across childhood (DuPaul, Power, Anastopoulos, & Reid, 1998; Luby et al., 2002; Weems & Costa, 2005).

The SDQ, originally developed in Great Britain and studied extensively in Europe, is unique in that it has separate parent, teacher, and youth self-report versions, and it has scales for assessing both symptoms and impairments (Bourdon, Goodman, Rae, Simpson, & Koretz, 2005). The SDQ was designed to assess five dimensions (conduct problems, peer problems, emotional symptoms, inattention/hyperactivity, and prosocial behaviors). However, exploratory and confirmatory factor analyses of the SDQ conducted with a U.S. sample showed that a three-factor solution (internalizing, externalizing, and prosocial behavior) provided the best fit to the data (Dickey & Blumberg, 2004). A similar three-factor model was identified in a U.S. sample based on youth self-report (Ruchkin, Jones, Vermeiren, & Schwab-Stone, 2008). The failure to distill a distinguishable factor related to ADHD, a disorder commonly assessed by providers in pediatric practice, seems to be a noteworthy limitation of the SDQ, at least with a U.S. population. Also, although the SDQ includes separate forms for younger (4–10 years) and older (11–17 years) children, differences between these versions are relatively minor and may not reflect important developmental variations between these age-groups. Furthermore, the predictive validity of the SDQ has been questioned when only one informant is used, which raises concerns about its feasibility for use in a primary care setting (Goodman, Ford, Simmons, Gatward, & Meltzer, 2003).

To address some of the limitations of existing screening measures, the Behavioral Health Checklist (BHCL) was developed. The BHCL is a parent-report measure designed to screen for common mental health conditions in the context of primary care. The measure was designed to: (1) provide valid and clinically meaningful screening of internalizing problems (anxiety and depression), externalizing problems (aggression and conduct problems), and ADHD (inattention and hyperactivity); (2) promote developmentally sensitive assessment across the age-range from 4 to 12 years; and (3) promote culturally sensitive assessment for the purposes of screening across urban and suburban populations, in particular white/Caucasian and black/African American children from low- and middle-income families.

The first goal of the study was to evaluate whether the proposed three-factor model provided an adequate fit to the data for two age-groups, 4–7 years and 8–12 years. The second goal was to evaluate model fit across the following demographic parameters for each age-group: Boys versus girls, urban versus suburban, white/Caucasian versus black/African American, and lower versus higher socioeconomic status, as assessed by subsidized lunch status.

Method

Characteristics of the Primary Care Practices

The data were collected in two urban primary care practices and three suburban practices affiliated with a tertiary-care children's hospital in a large metropolitan area in the northeast section of the United States. One urban practice

served about 9,500 patients per year and the other served about 10,000 per year. In the urban practices, approximately 75% of the children were eligible for publicly funded medical assistance; 89% were non-Hispanic, 3% were Hispanic, and 8% were of unknown ethnicity. With regard to race for the urban sample, 91% were black/African American, 2.5% were white/Caucasian, 1% were Asian, 0.5% were other races, and 5% were unknown.

The three suburban practices served between 4,200 and 6,900 patients per year. In the suburban practices, approximately 20% of the children were eligible for medical assistance; 91% were non-Hispanic, 2% were Hispanic, and 7% were of unknown ethnicity. With regard to race for the suburban sample, 62% were white/Caucasian, 23% were black/African American, 3% were Asian, 1% were other races, and 11% were unknown.

Recruitment Procedures

Parents were approached by trained research staff while waiting in the primary care office waiting room for an appointment. Parents were eligible if their children met the age criteria for the study (4–12 years) and they were able to read or understand English. Parents who could not read English because it was not their first language (less than five parents) were asked to respond orally to questions read to them. At the outset of the study, parents with multiple children were asked to complete measures for the oldest child in the study age-range. During the latter stages of the study, parents with multiple children were asked to select the child belonging to age-groups that were relatively underrepresented in the sample.

Between December 2010 and September 2012, 2,171 parents were approached to participate in the study. Of these parents, 1,702 (78.4%) met eligibility criteria. The primary reason for ineligibility was that the parents did not have a child in the study age-range (about 95% of ineligible cases); other reasons were that the primary caregivers were not present during the screening, the parents had been previously invited to participate in the study, and the parents did not understand English. Of the eligible cases, 1,406 (82.6%) consented to participate. The number of boys and girls for whom measures were completed in the urban and suburban practices for each age-group (4–7 and 8–12 years) is presented in Table I. Because we did not collect any identifying information from participants, this study was determined to be exempt by the hospital's Institutional Review Board. Given grant funding available to support this project (separate grants for data collection in the urban and suburban sites), parents from the urban practices received a stipend of $20 for participation and those from suburban practices were offered a stipend of $10.

Table I. *Number of Participants From Urban and Suburban Practices for Each Age-Group*

Age group	Female	Male	All children
Urban practices			
4–7 years	280	232	512
8–12 years	176	162	338
Total urban	456	394	850
Suburban practices			
4–7 years	147	132	279
8–12 years	118	148	266
Total suburban	265	280	545
Total urban and suburban	721	674	1,395

Note. The table indicates those participants who were included in the analyses.

Measures

Participants completed a demographic form as well as the BHCL. In addition, most participants (94% for the younger age-group and 84% for the older group) completed the Child Behavior Checklist (CBCL) to evaluate concurrent validity.

Demographic Form

Parents were asked to complete a brief demographic questionnaire indicating (1) their relationship with the child, (2) the child's age, (3) the child's gender, (4) the child's race and ethnicity, (5) the primary caregiver's highest level of education, (6) the family's annual household income, (7) the number of people in the household, (8) whether the child was eligible for Head Start, (9) whether the child was eligible for free or reduced-price school lunch, and (10) the primary language spoken in the home.

Behavioral Health Checklist

The BHCL is a 27-item parent-report measure for screening for behavioral health problems among children presenting to a primary care practice, including 18 items assessing emotional and behavioral problems and nine items assessing adaptive functioning to promote a strengths-based assessment (adaptive items are analyzed in a separate study). Items are rated on a 4-point scale: 0 = never or rarely, 1 = sometimes, 2 = often, and 3 = very often. The BHCL was developed to include six items related to internalizing problems (three anxiety and three depression), six items pertaining to externalizing problems (three aggression and three oppositional/conduct problems), and six items referring to inattention/hyperactivity (three inattention and three hyperactivity). Separate versions were created for younger (4–7 years) versus older (8–12 years) children, which generally corresponds to the periods of early versus middle childhood (U.S. Department of Health and

Human Services, 2013). The measure generally takes less than 10 min to complete.

The BHCL was developed after careful review of existing standardized instruments and research on behavioral assessment. Items were selected based on demonstration of adequate and unique factor loadings on one of three domains (internalizing problems, externalizing problems, or inattention/hyperactivity). Specifically, items selected for consideration generally were those with the highest factor loadings on the target factor that did not load on other factors based on existing empirical evidence [e.g., PSC-17; Behavior Assessment System for Children – Second Edition (Reynolds & Kamphaus, 2004); CBCL (Achenbach & Rescorla, 2001a); and ADHD Rating Scale – IV (DuPaul et al., 1998)]. In addition, items with the highest total predictive power for assessing disorders were given serious consideration (Power, Costigan, Leff, Eiraldi, & Landau, 2001).

Given developmental differences in the expression of emotional and behavioral symptoms (Achenbach & Rescorla, 2001a, b; Luby et al., 2002; Reynolds & Kamphaus, 2004; Weems & Costa, 2005), distinct item sets were created to screen younger (4–7 years) and older (8–12 years) children. For example, an Inattention/Hyperactivity item on the 4–7-year version was worded ''Shifts activities too quickly,'' whereas a similar item on the 8–12-year version was worded ''Has difficulty sustaining attention.'' Also, on the Externalizing factor, an item that was included only on the 4–7-year version was ''Refuses to share,'' whereas an item that was included only on the 8–12-year version was ''Lies or cheats.'' Further, on the Internalizing factor, an item included only on the 4–7-year version was ''Is easily annoyed or cranky,'' whereas an item included only on the 8–12-year version was ''Is down on self.'' The wording of items was modified to ensure brevity as well as consistent tense and grammatical style. After identifying the initial set of items, 16 parents of diverse racial, ethnic, and socioeconomic backgrounds were interviewed regarding the understandability, appropriateness, and readability of items contained in both versions of the BHCL. Only minor formatting revisions (e.g., spacing, font size) were suggested, and these changes were included in the final versions of the measure.

Child Behavior Checklist

The CBCL version for 6–18-year-old children Achenbach & Rescorla (2001a) was administered to parents of children between 6 and 12 years; and the CBCL version for 1.5–5-year-old children (Achenbach & Rescorla, 2001b) was administered to parents of children who were 4 and 5 years of age. Both versions of the CBCL yield broad-band Internalizing and Externalizing dimensions that correspond to the Internalizing and Externalizing factors of the BHCL. Also, both versions of the CBCL contain an Attention Problems factor that assesses a dimension corresponding to the Inattention/Hyperactivity factor of the BHCL. The CBCL was selected for the examination of concurrent validity because it is widely regarded as a benchmark for the empirical assessment of child psychopathology; hundreds of studies attest to the validity and clinical utility of this measurement system.

Data Analyses

The data were subjected to confirmatory factor analysis to test the theoretically specified structure against a priori alternative models (MacCallum, 2003). Given the strong theoretical foundation of the BHCL, a three-factor solution was expected. For the younger version of the BHCL (see Table IV for item descriptions), our expectation was that Items 1, 5, 8, 11, 13, and 15 would load on an Internalizing factor; Items 3, 7, 9, 12, 16, and 17 on an Externalizing factor; and Items 2, 4, 6, 10, 14, and 18 on an Inattention/Hyperactivity factor. For the older version of the scale (see Table VI for item descriptions), our belief was that Items 1, 5, 7, 9, 11, and 17 would load on an Internalizing factor; Items 3, 8, 12, 13, 15, and 18 on an Externalizing factor; and Items 2, 4, 6, 10, 14, and 16 on an Inattention/Hyperactivity factor.

Analyses were conducted with Mplus version 7.1 for Macintosh. Given the non-normality of the data, items were treated as categorical indicators of continuous factors using the WLSMV estimator (Flora & Curran, 2004). Although there is no consensus on the proper cutoff values for fit indices (Marsh, Hau, & Wen, 2004) and characteristics of the model and data can affect the performance of fit statistics (West, Taylor, & Wu, 2012), comparative fit index (CFI) values $\geq$.90 and root mean square error of approximation (RMSEA) values$\leq$.08 may indicate acceptable fit, whereas CFI values $\geq$.95 and RMSEA values $\leq$.05 may reflect a good fit to the data (Marsh, Liem, Martin, Morin, & Nagengast, 2011). In agreement with these rules of thumb, Hu and Bentler (1998, 1999) recommended a dual cutoff value of .95 for the CFI and .06 for the RMSEA to ensure against both Type I and Type II errors.

Given these guidelines, models with CFI $\geq$.95 and RMSEA $\leq$.06 were considered good; models with CFI $\geq$.90 and RMSEA $\leq$.08 were deemed acceptable; and models with the upper limit of 90% confidence intervals for RMSEA $>$.10 were deemed unacceptable (Browne & Cudeck, 1993).

Given the excessive power associated with large sample sizes and the large number of model comparisons, the difference between models was evaluated using practical fit criteria rather than a statistical test of the differences in chi-square values (Gignac, 2007). Criteria considered were a change in CFI values $\geq$.01 and a change in RMSEA values $\geq$.02, which are indicative of a meaningful improvement in model fit (Chen, 2007; Cheung & Rensvold, 2002). A second consideration was whether the model fit for each of the subgroups of the sample (urban vs. suburban, older vs. younger children in the age-group, boys vs. girls, African American vs. other races, and subsidized vs. unsubsidized lunch status [as a proxy for socioeconomic status]) was adequate. Even though most of the children in the other race (not African American) subgroup were white/Caucasian, separate analyses were conducted for the white subsample. Given that the intent of scientific research is to find reproducible results (Haig, 2005; Rummel, 1967), a model that replicates across samples is an important goal (Preacher, Zhang, Kim, & Mels, 2013). Accordingly, the generalizability of model fit across subgroups was evaluated according to the CFI and RMSEA standards described previously for good, acceptable, and unacceptable fit. Although the overall sample size in this study was relatively large, the size of the subsamples was generally unbalanced and inadequate for statistical tests of invariance across subgroups (Meade & Bauer, 2007).

As a final step, Pearson correlations were conducted to examine the relationship of BHCL factors to comparable factors of the CBCL (Internalizing, Externalizing, and Attention Problems). Separate correlational analyses were conducted for the 4–7-year-old and 8–12-year-old versions of the BHCL.

Results

The findings are reported separately for the younger and older versions. Table II provides information about the demographic characteristics of children from the urban and suburban practices whose caregivers participated in this study. As expected, there were significant differences between the urban and suburban practices for each age-group with regard to race and eligibility for subsidized lunch. Children in the urban and suburban practices did not differ significantly with regard to age and gender distribution within each of the two age-groups.

Factor Analyses of the Younger (4–7-Year) Version

For the younger version, there were 794 cases in total. Three cases with more than three missing items were discarded. Of the remaining 791 cases, 56 were missing one item and 9 were missing two items. These 0.52% missing data points were imputed with the EM algorithm of SPSS version 19 for the Macintosh. As expected based on the use of a nonclinical sample, the responses were non-normally distributed. Univariate skew ranged from 0.02 to 2.28 and univariate kurtosis ranged from −0.11 to 6.45. Mardia's multivariate skew and kurtosis (Mardia, 1970) were 56.23 and 88.25, respectively.

The three-factor theoretical model with Internalizing Problems, Externalizing Problems, and Inattention/Hyperactivity was contrasted with a one-factor model of general psychopathology and a two-factor model composed of an Internalizing factor and an Externalizing factor (combining items pertaining to Externalizing and Inattention/Hyperactivity). Results are presented in Table III. Of these three models, the theoretical model clearly was the best fit

Table II. *Demographic Characteristics of Participants From the Urban and Suburban Practices for the 4–7-Year and 8–12-Year Versions of the BHCL*

	4–7 years			8–12 years		
Demographic factors	Urban	Suburban	*p*	Urban	Suburban	*p*
Age in years – mean (*SD*)	5.6 (1.1)	5.4 (1.2)	ns	9.7 (1.4)	9.9 (1.4)	ns
Gender – female (%)	54.7	52.7	ns	52.1	44.4	ns
Black/African American (%)	89.8	18.0	**	89.3	19.0	**
Asian (%)	2.2	3.6		1.8	2.3	
White/Caucasian (%)	1.0	69.4		1.2	71.9	
Other race (%)	7.0	9.0		7.7	6.8	
Latino/Hispanic (%)	4.5	6.9	ns	2.5	6.3	*
Eligible for subsidized lunch (%)	81.0	27.9	**	78.2	29.7	**

Note. *t*-tests were used to examine age differences between the urban and suburban samples within each age-group. Chi-square tests were used to examine group differences in gender, race, ethnicity, and eligibility for subsidized lunch.

ns indicates that the urban and suburban groups did not differ within that age-group.

*$p < .05$, **$p < .001$.

Table III. *Fit of Models to BHCL 4–7 Data for Total and Subsamples*

Sample/model	*N*	χ^2	*df*	CFI	RMSEA	RMSEA 90% CI
Total sample						
One-factor	791	1,553.92	135	.87	.12	.11–.12
Two-factor	791	1,203.01	134	.90	.10	.10–.11
Theoretical	791	825.41	132	.94	.08	.08–.09
Subsamples (theoretical)						
Urban	512	591.88	132	.93	.08	.08–.09
Suburban	279	374.38	132	.95	.08	.07–.09
Ages 4–5	388	417.10	132	.95	.08	.07–.08
Ages 6–7	403	478.48	132	.94	.08	.07–.09
Boys	364	494.66	132	.93	.09	.08–.10
Girls	427	470.51	132	.94	.08	.07–.09
African American	507	580.98	132	.93	.08	.08–.09
Not African American	280	336.96	132	.96	.07	.07–.08
White	198	332.33	132	.97	.09	.08–.10
Subsidized lunch	473	505.73	132	.94	.08	.07–.09
No subsidized lunch	289	446.18	132	.92	.09	.08–.10

Table IV. *Standardized Coefficients for BHCL 4–7 (N = 791) Theoretical Model*

Item	Internal	External	ADHD
1. Feels sad	.646		
2. Shifts activities too quickly			.672
3. Fights with other children		.781	
4. Has poor self-control			.837
5. Afraid of making mistakes	.390		
6. Is unable to slow down			.771
7. Argues with adults		.703	
8. Worries	.516		
9. Is defiant		.844	
10. Can't sit still			.827
11. Is easily annoyed or cranky	.871		
12. Refuses to share		.636	
13. Is too fearful	.579		
14. Has trouble paying attention			.887
15. Is unhappy	.793		
16. Is disobedient		.866	
17. Argues with peers		.751	
18. Can't concentrate			.831
Alpha (total = .901)	.704	.836	.862

to the data (ΔCFI = .04 and ΔRMSEA = .02 when compared with the two-factor model), resulting in acceptable fit values for CFI and RMSEA. Accordingly, the theoretical model was accepted as the best explanation of these data. Standardized loadings for the total sample are presented in Table IV.

To examine generalizability, the theoretical model was tested in various subsamples of children. Results from these analyses are reported in Table III. All subsamples generally demonstrated acceptable fit; although for boys, students without subsidized lunch, and children who were white/Caucasian, the RMSEA value just failed to reach the criterion for acceptability. However, none of the subsamples demonstrated unacceptable fit.

Alpha coefficients for Internalizing Problems, Externalizing Problems, and Inattention/Hyperactivity scales were .70, .84, and .86, respectively, which are all in the acceptable range. The total score reliability was .90. Scores on the Internalizing factor correlated .52 with the Externalizing factor and .51 with the Inattention/Hyperactivity factor. Scores on the Externalizing factor correlated .66 with the Inattention/Hyperactivity factor. Means and standard deviations for each factor of the 4–7-year version were as follows: Internalizing ($M = 8.80$, $SD = 2.44$), Externalizing ($M = 11.88$, $SD = 4.06$), Inattention/Hyperactivity ($M = 10.04$, $SD = 3.82$).

Factor Analyses of the Older (8–12-Year) Version

There were 604 cases in total, with 40 missing data points (0.38%) that were imputed with the EM algorithm of SPSS version 19 for the Macintosh. As expected, the responses were non-normally distributed. Univariate skew ranged from 0.48 to 4.11 and univariate kurtosis ranged from −0.84 to 18.44. Mardia's multivariate skew and kurtosis (Mardia, 1970) were 73.69 and 78.17, respectively.

Like the version for the young children, the three-factor theoretical model with Externalizing Problems, Internalizing Problems, and Inattention/Hyperactivity factors was contrasted with a one-factor model of general psychopathology and a two-factor model composed of Internalizing and Externalizing factors. Results are presented in Table V. Of these three models, the theoretical model with Internalizing Problems, Externalizing Problems, and Inattention/Hyperactivity factors was a good fit to the data and superior to the alternative models (ΔCFI = .04, ΔRMSEA = .03 when compared with the two-factor model). As with the younger age-group, the theoretical model was accepted as the best explanation of these data. Standardized coefficients for that model are presented in Table VI. To examine generalizability, the theoretical model was tested in various subsamples of children (Table V). The results of these analyses demonstrated that the theoretical model produced acceptable to good fit in all subgroups.

Alpha coefficients for Internalizing Problems, Externalizing Problems, and Inattention/Hyperactivity were .84, .82, and .84, respectively, which are well above

Table V. *Fit of Models to BHCL 8–12 Data for Total and Subsamples*

Sample/model	*N*	χ^2	*df*	CFI	RMSEA	RMSEA 90% CI
Total sample						
One-factor	604	1,527.97	135	.83	.13	.13–.14
Two-factor	604	767.53	134	.92	.09	.08–.10
Theoretical	604	442.80	132	.96	.06	.06–.07
Subsamples (theoretical)						
Urban	338	309.96	132	.96	.06	.05–.07
Suburban	266	331.68	132	.94	.08	.07–.09
Grades <5	352	292.46	132	.97	.06	.05–.07
Grades 5–8	248	266.99	132	.96	.06	.05–.08
Boys	310	264.28	132	.98	.06	.05–.07
Girls	294	274.03	132	.96	.06	.05–.07
African American	351	266.93	132	.97	.05	.05–.06
Not African American	249	343.99	132	.93	.08	.07–.09
White	193	288.50	132	.92	.08	.07–.09
Subsidized lunch	337	281.95	132	.97	.06	.05–.07
Nonsubsidized lunch	258	276.67	132	.95	.07	.05–.08

Table VI. *Standardized Coefficients for BHCL 8–12 (N = 604) Theoretical Model*

Item	Internal	External	ADHD
1. Feels hopeless	.854		
2. Has difficulty sustaining attention			.806
3. Fights with other children		.802	
4. Talks excessively			.625
5. Is afraid of making mistakes	.641		
6. Has difficulty awaiting turn			.799
7. Is down on self	.847		
8. Steals		.618	
9. Worries	.777		
10. Fidgets or squirms in seat			.788
11. Feels sad	.854		
12. Teases others		.820	
13. Loses temper too easily		.833	
14. Avoids tasks that require effort			.786
15. Breaks the rules		.841	
16. Loses things necessary for activities			.756
17. Is too fearful	.734		
18. Lies or cheats		.699	
Alpha (total = .897)	.835	.822	.840

the acceptable range. The total score reliability was .90. Scores on the Internalizing factor correlated .36 with the Externalizing factor and .53 with the Inattention/Hyperactivity factor. Scores on the Externalizing factor correlated .62 with the Inattention/Hyperactivity factor. Means and standard deviations for each factor of the 8–12-year version were as follows: Internalizing ($M = 9.23$, $SD = 3.07$), Externalizing ($M = 8.80$, $SD = 3.07$), Inattention/Hyperactivity ($M = 10.59$, $SD = 3.97$).

Table VII. *Pearson Correlations Between BHCL Subscales and Corresponding CBCL Empirical Scales*

CBCL scale	BHCL Internalizing	BHCL Externalizing	BHCL Inattention/ Hyperactivity
4–7-year-old children			
CBCL Internalizing Problems	.64	.51	.50
CBCL Externalizing Problems	.55	.76	.71
CBCL Attention Problems	.51	.58	.79
8–12-year-olds			
CBCL Internalizing Problems	.68	.35	.50
CBCL Externalizing Problems	.45	.77	.68
CBCL Attention Problems	.42	.45	.75

Note. All correlations are significant at the .001 level.

Correlations With the CBCL

Correlations between the BHCL and comparable factors of the CBCL are presented in Table VII. For every correlation computed, the BHCL factor had a higher correlation with its corresponding factor on the CBCL than with noncorresponding factors. In general, the pattern of divergence with noncorresponding factors was greater for the older than the younger version of the BHCL, with the exception of the Inattention/Hyperactivity factor of the BHCL, which differed in its correlation with the CBCL Attention Problems and Externalizing factors by .08 for the younger children and .07 for the older group.

Discussion

The findings of this study support the construct validity of the BHCL for children across the age-range from 4 to 12 years. The hypothesized three-factor structure of the BHCL, including factors pertaining to Internalizing Problems, Externalizing Problems, and Inattention/Hyperactivity, was supported. For both age-groups studied (4–7 years and 8–12 years), the theoretical model exhibited acceptable to good fit to the data for the total sample and for each subsample studied. The pattern of correlations between the BHCL and corresponding and noncorresponding factors of the CBCL offers further evidence of the validity of the BHCL.

The BHCL represents an advancement in behavioral health screening, in that the factors of this scale are aligned with dimensions of child functioning that are both theoretically and clinically meaningful. Of note, the BHCL Internalizing factor contains an equal number of items assessing both anxiety (three items) and depression

(three items). In contrast, the PSC-17 includes only one item that clearly maps to anxiety (i.e., ''worries a lot''; Stoppelbein et al., 2012). Further, the BHCL includes a clearly distinguishable factor pertaining to ADHD within the targeted age-group, whereas the SDQ may not (Dickey & Blumberg, 2004).

Another asset of the BHCL is its developmental appropriateness across the 4–12 age-range. Each version contains a set of items that have been shown in previous research to be appropriate and empirically supported for the targeted age-group (Achenbach & Rescorla, 2001a, b; Reynolds & Kamphaus, 2004). The findings of this study indicate that model fit for children in the lower and upper age/grade-level subgroups of both versions of the BHCL (i.e., ages 4–5 and 6–7 for the younger version, and grades <5 and grades 5–8 for the older version) is at least adequate. In addition, the three-factor BHCL model was demonstrated to have acceptable fit across a broad range of demographic groups. Model fit was adequate across urban and suburban practices. In addition, the model was applicable for girls and boys, children of lower and middle socioeconomic status, and children who were black/African American and white/Caucasian. Applying the criteria for meaningful differences between factor models, there was evidence that model fit was stronger for the urban (as compared with suburban) subsample and the black/African American (compared with the non-African American and Caucasian) subgroup.

The BHCL factors identified in this study focus solely on children's deficits. These factors fail to acknowledge children's adaptive functioning and the potential for children's strengths to serve as protective factors with regard to social and academic functioning (Kwon, Kim, & Sheridan, 2012; Suldo & Shaffer, 2008). The BHCL includes a set of adaptive items that were not analyzed in this study. In future research, it will be important to determine whether these items form a cohesive factor, whether this scale moderates (i.e., serves as a protective factor in) the relationship between dimensions of psychopathology and functional impairments, and whether these items have clinical utility for screening in primary care.

The study has several limitations that deserve mention. Several of the limitations pertain to the generalizability of the findings. First and perhaps most significantly, although the sample was highly diverse, children of Hispanic ethnicity and those of minority races other than black/African American were unrepresented in the current study. As a result, the applicability of the three-factor model to these populations is not known. Second, the black/African sample was disproportionately included in the lower socioeconomic group and underrepresented in the middle socioeconomic group. The sample was too small to statistically test differences as a function of both race and socioeconomic status (i.e., black/African American children eligible and ineligible for subsidized lunch vs. white/Caucasian children eligible and ineligible for subsidized lunch). As such, the sample in this study may not represent black/African American and white/Caucasian families of both lower and middle socioeconomic status. Third, the study was conducted with families in urban and suburban communities located in a large metropolitan area in the northeast section of the United States. The applicability of the findings to rural areas and populations outside the northeast needs to be demonstrated. Fourth, the BHCL was developed for the screening of children between the age-ranges of 4 and 12 years. Alternative screening methods are required for the screening of children older than 12. Fifth, although this study provides considerable support for the construct validity of the BHCL, particularly with regard to the generalizability of the factor structure and pattern of item loadings across groups, additional research is needed to demonstrate metric invariance (equivalence of the magnitude of factor loadings across groups) and scalar invariance (equivalence of item intercepts across groups), as described in a recent study conducted by Stoppelbein et al. (2012). Finally, additional research is needed to demonstrate the predictive validity and clinical utility of the BHCL when used for screening in pediatric primary care practices. Although the BHCL may have some advantages when compared with the PSC-17 and SDQ, a comparative analysis of these measures would be useful.

In conclusion, this study provides substantial evidence of the construct validity of a new measure for screening behavioral health problems in primary care. The BHCL was shown to be developmentally appropriate across the age-range from 4 to 12 years. This study provides evidence of the cultural sensitivity of the measure across black/African American and white/Caucasian families as well as those of lower and middle income groups. However, Hispanic families and those from other races were underrepresented in this study. The BHCL addresses many of the limitations of current screening tools, but additional research regarding the cultural sensitivity as well as construct and predictive validity of this measure are needed. In addition, norms derived from a large representative sample across the United States are needed for the BHCL to be useful in clinical settings.

Funding

This study was supported by the Foerderer Award, granted by the Research Institute of The Children's Hospital of

Philadelphia, and the Academic Pediatric Association/ Bright Futures Young Investigator Award, granted to Anson Koshy. Also, the study was supported in part by Projects T73MC00051 and T77MC00012 from the Maternal and Child Health Bureau (Public Health Service Act, Section 399BB (e) (1) (A), as amended by the Combating Autism Act of 2006), Health Resources and Services Administration, Department of Health and Human Services. In addition, The Children's Hospital of Philadelphia received support from Shire for fellowship training, which partially supported this project.

Conflicts of interest: Thomas Power derives royalties from Guilford Press for the ADHD Rating Scale - IV, a measure referenced in this article.

References

Achenbach, T. M., & Rescorla, L. A. (2001a). *Manual for the ASEBA school-age forms and profiles: An integrated system of multi-informant assessment*. Burlington, VA: Library of Congress.

Achenbach, T. M., & Rescorla, L. A. (2001b). *Manual for the ASEBA preschool forms and profiles*. Burlington, VA: Library of Congress.

American Academic of Pediatrics Task Force on Mental Health. (2010). The case for routine mental health screening. *Pediatrics, 125*, S133–S139.

Bell, J. F., Johnson, M. L., Myers, K., & Patrick, D. L. (2010). Primary health care quality in a national sample of children and youth with mental health impairment. *Journal of Developmental and Behavioral Pediatrics, 31*, 694–704.

Bourdon, K. H., Goodman, R., Rae, D. S., Simpson, G., & Koretz, D. S. (2005). The strengths and difficulties questionnaire: U.S. normative data and psychometric properties. *Journal American Academy of Child Adolescent Psychaitry, 44*, 557–664.

Braveman, P., & Barclay, C. (2009). Health disparities beginning in childhood: A life-course perspective. *Pediatrics, 124(Suppl 3),* S163–S175.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464–504.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233–255.

Dickey, W. C., & Blumberg, S. J. (2004). Revisiting the factor structure of the strengths and difficulties questionnaire: United States, 2001. *Journal of the American Academy of Child & Adolescent Psychiatry, 43*, 1159–1167.

DuPaul, G. J., Power, T. J., Anastopoulos, A. D., & Reid, R. (1998). *The ADHD Rating Scale–IV: Checklists, norms, and clinical interpretation*. New York, NY: Guilford Press.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*, 466–491.

Flores, G., & The Committee on Pediatric Research. (2010). Racial and ethnic disparities in the health and health care of children. *Pediatrics, 125*(4), e979–e1020.

Foy, J. M., Kelleher, K. J., & Laraque, D., & American Academy of Pediatrics Task Force on Mental Health. (2010). Enhancing pediatric mental health care: Algorithms for primary care. *Pediatrics, 125*, S109–S125. doi:10.1542/peds.2010-0788F

Gardner, W., Lucas, A., Kolko, D. J., & Campo, J. V. (2007). Comparison of the PSC-17 and alternative mental health screens in an at-risk primary care sample. *Journal of the American Academy of Child and Adolescent Psychiatry, 46*, 611–618.

Gardner, W., Murphy, M., Childs, G., Kelleher, K., Pagano, M., Jellinek, M. S., . . . Chiappetta, L. (1999). The PSC-17: a brief pediatric symptom checklist with psychosocialproblem subscales. A report from PROS and ASPN. *Ambulatory Child Health, 5*, 225–236.

Gignac, G. E. (2007). Multi-factor modeling in individual differences research: Some recommendations and suggestions. *Personality and Individual Differences, 42*, 37–48.

Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2003). Using the strengths and difficulties questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *International Review of Psychiatry, 15*, 166–172. doi:10.1080/0954026021000046128

Haig, B. D. (2005). An abductive theory of scientific method. *Psychological Method, 10*, 371–388. doi:10.1037/1082-989X.10.4.371

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to under parameterized model misspecification. *Psychological Methods, 3*, 424–453.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55. doi:10.1080/10705519909540118

Kostanecka, A., Power, T. J., Clarke, A., Watkins, M., Hausman, C. L., & Blum, N. J. (2008). Behavioral health screening in urban primary care settings: Construct validity of the PSC-17. *Journal of Developmental and Behavioral Pediatrics, 29*, 124–128. doi:10.1097/DBP.0b013e31816a0d9e

Kwon, K., Kim, E., & Sheridan, S. (2012). Behavioral competence and academic functioning among early elementary children with externalizing problems. *School Psychology Review, 41*, 123–140.

Luby, J. L., Heffelfinger, A. K., Mrakotsky, C., Hessler, M. J., Brown, K. M., & Hildebrand, T. (2002). Preschool Major Depressive Disorder: Preliminary validation for developmentally modified DSM-IV Criteria. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*, 928–936.

MacCallum, R. C. (2003). Working with imperfect models. *Behavioral Research, 38*, 113–139.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57*, 519–530.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320–341.

Marsh, H. W., Liem, G. A. D., Martin, A. J., Morin, A. J. S., & Nagengast, B. (2011). Methodological measurement fruitfulness of exploratory structural equation modeling (ESEM): New approaches to key substantive issues in motivation and engagement. *Journal of Psychoeducational Assessment, 29*, 322–346.

Meade, A. W., & Bauer, D. J. (2007). Power and precision in confirmatory factor analytic tests of measurement invariance. *Structural Equation Modeling, 14*, 611–635.

Navon, M., Nelson, D., Pagano, M., & Murphy, M. (2001). Use of the pediatric symptom checklist in strategies to improve preventive behavioral health care. *Psychiatric Services, 52*, 800–804.

Power, T. J., Costigan, T. E., Leff, S. S., Eiraldi, R. B., & Landau, S. (2001). Assessing ADHD across settings: Contributions of behavioral assessment to categorical decision making. *Journal of Clinical Child Psychology, 30*, 399–412.

Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing an optimal number of factors in exploratory factor analysis: A model selection perspective. *Multivariate Behavioral Research, 48*, 28–56. doi:10.1080/00273171.2012.710386

Reynolds, C. R., & Kamphaus, R. W. (2004). *BASC-2: Behavior Assessment System for Children* (2nd ed.). Circle Pines, MN: AGS Publishing.

Ruchkin, V., Jones, S., Vermeiren, R., & Schwab-Stone, M. (2008). The strengths and difficulties questionnaire: The self-report version in American urban and suburban youth. *Psychological Assessment, 20*(2), 175–82. doi:10.1037/1040-3590.20.2.175

Rummel, R. J. (1967). Understanding factor analysis. *Journal of Conflict Resolution, 11*, 444–480.

Stoppelbein, L., Greening, L., Moll, G., Jordan, S., & Suozzi, A. (2012). Factor analysis of the pediatric symptom checklist-17 with African-American and Caucasian pediatric populations. *Journal of Pediatric Psychology, 37*, 348–357.

Suldo, S. M., & Shaffer, E. J. (2008). Looking beyond psychopathology: The Dual-Factor Model of mental health in youth. *School Psychology Review, 37*, 52–68.

U.S. Department of Health and Human Services. (2013). Healthy People 2020. Retrieved from www.healthypeople.gov/2020/default/aspz. Retrieved June 20, 2013.

Van Cleave, J., Gortmaker, S. L., & Perrin, J. M. (2010). Dynamics of obesity and chronic health conditions among children and youth. *JAMA-Journal of the American Medical Association, 303*, 623–630.

Weems, C. F., & Costa, N. M. (2005). Developmental differences in the expression of childhood anxiety symptoms and fears. *Journal of the American Academy of Child and Adolescent Psychiatry, 44*, 656–663.

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209–231). New York, NY: Guilford.