Article

Validating the Pediatric Symptoms Checklist–17 in the Preschool Environment

Journal of Psychoeducational Assessment
2020, Vol. 38(4) 460–474

DOI: 10.1177/0734282919828234
journals.sagepub.com/home/jpa
SAGE

Jin Liu[1], Yin Burgess[1], Christine DiStefano[1], Fan Pan[1], and Ning Jiang[1]

Abstract

In the Response to Intervention framework, a psychometrically sound screening tool is essential for identification of children with emotional and behavioral risk. The purpose of this study was to examine the validity of the Pediatric Symptom Checklist–17 (PSC-17) screener in school-based settings. Forty-four teachers rated 738 preschoolers using the PSC-17; children were later assessed using long forms of the Behavior Assessment System for Children (BASC-2) Preschool form or the Achenbach System of Empirically Based Assessment (ASEBA) Caregiver–Teacher Report Form to identify emotional and behavioral disorder. Validity evidence including examinations of a multilevel factor structure, internal consistency, and criterion-related validity supported the conclusion that the PSC-17 is a high-quality universal screening tool in school-based settings. Finally, to identify emotional and behavioral risk with young children, receiver operating characteristic curve analyses with the PSC-17 yielded a lower cutoff score (i.e., 7) than the original cutoff score (i.e., 15) based on a clinical sample.

Keywords

PSC-17, young children, validation, teacher rating, cutoff scores

For well over a decade, research studies have noted adverse effects of emotional and behavioral disorders (EBD) on educational and social outcomes. For example, the presence of EBD has been found to be related to negative outcomes such as academic achievement, school dropout, teenage pregnancy, unemployment, and youth delinquency (Bullis & Yovanoff, 2006; Durbrow, Schaefer, & Jimerson, 2000; Gutman, Sameroff, & Cole, 2003; Jones, Greenberg, & Crowley, 2015; Lane, Little, Menzies, Lambert, & Wehby, 2010; McEvoy & Welker, 2000; A. J. Reynolds, Temple, Robertson, & Mann, 2001). Given the possible detrimental effects associated with EBD, parents, teachers, and school psychologists are interested in identifying these disorders in children as soon as possible.

Many children are detected with EBD in the school setting. Schools typically have used a "refer-test-place" model to identify children with EBD, which relies on teachers to identify

[1]University of South Carolina, Columbia, SC, USA

Corresponding Author:
Jin Liu, Department of Educational Studies, College of Education, University of South Carolina, 145 Wardlaw, Columbia, SC 29208, USA.
Email: liu99@mailbox.sc.edu

children exhibiting problematic behaviors and emotions in the classroom as well as on school psychologists to determine the eligibility of special education based on assessment results (DiStefano & Morgan, 2011). Waiting to collect evidence for referral allows extended time for problems to become ingrained and results in increased health care expenses due to later treatment costs (Aos, Lieb, Mayfield, Miller, & Pennucci, 2004).

Alternatively, researchers have brought forth the idea of early intervention to minimize the adverse effects of EBD. The Response to Intervention framework (RtI) has been proposed as an alternative (Ikeda, Neessen, & Witt, 2008). Recently, the RtI framework has been extended to the early childhood level (Carta & Greenwood, 2013); however, RtI in preschool must be adapted to fit within the early childhood setting and to address the needs of young children (Coleman, Roth, & West, 2009; Greenwood et al., 2011). The RtI framework typically includes three tiers of assistance. The first tier includes an evidence-based curriculum and intentional teaching that promote child development and school readiness to prevent, identify, and delay the onset of developmental, academic, and behavioral problems. Universal screening is typically used to identify those at-risk children who may need additional support. In Tier 2, children identified as at risk typically receive supplemental instruction in a general classroom (e.g., small group intervention). Progress monitoring is conducted to guide interventions with involvement of parents and family members. In Tier 3, children who are not making adequate progress in Tier 2 may be referred for comprehensive assessment and more intensive intervention assistance. More frequent program monitoring is used to guide decisions of a child's final status. The overall goal of RtI is to implement a systematic approach to lessen the negative outcomes.

Preschool is likely to be the first school setting where young children may be eligible to receive RtI assistance. As preschool teachers have opportunities to observe a large number of young children simultaneously (Levitt, Saka, Romanelli, & Hoagwood, 2007), RtI implemented by teachers may allow for the collection of consistent and reliable information about preschool children's emotional and behavioral risk (EBR) status. Thus, the current study focused on universal screening of EBR by teachers at school.

Universal screening plays an important role by providing an efficient way to examine if all children's development is on target and/or if additional assessment is required (Coleman et al., 2009; Greer & Liu, 2016). There are several barriers to early identification of EBR for young children. Screening tools often lack technical adequacies compared with comprehensive assessment measures, which rely on more extensive and rigorous research procedures (Gokiert et al., 2014). This could be due to the relative short history of behavioral screening for young children (Greenwood et al., 2011; Steed & Banerjee, 2016). For example, it is more difficult to measure internalizing problems for young children who have a limited capacity to express their internal feelings (Tandon, Cardeli, & Luby, 2009). In addition, professionals, such as teachers, are not well-trained to identify EBR and often cannot provide needed services for children at school (Hemmeter, Santos, & Ostrosky, 2008). Thus, identifying a high-quality screening tool that can be used for effective universal screening by preschool teachers is of great importance.

The PSC-17

The PSC-17 aims to "improve the recognition and treatment of psychosocial problems in children" (Massachusetts General Hospital, 2007-2017, "Pediatric Symptom Checklist," para. 1). The scale was originally developed by Gardner and colleagues (1999) by shortening the full PSC (i.e., a 35-item measure) using exploratory factor analysis (EFA). To provide support for the scale, two large primary care research networks were studied, with data provided by parents during a primary care visit. A sample of 18,045 children aged 4 to 15 years was included. Seventeen items were retained from the EFA with three important dimensions identified—Externalizing Problems, Internalizing Problems, and Attention Problems. Externalizing Problems measure the disruptive behaviors, such as aggression and hyperactivity (e.g., fights with other children); Internalizing Problems measure feelings of depression, worry, and anxiety (e.g., feels sad, unhappy); and Attention Problems measure attention-deficit issues (e.g., distracted easily). The PSC total score combines the three subscales to provide a measure of overall maladaptive behavior.

To identify children at risk, cutoff scores were developed with a sample of 406 children from a hospital-based mental health clinic. Those children were referred for psychosocial problems or received psychological service after their parents filled out the PSC-17. Previously validated instruments, Screen for Childhood Anxiety Related Emotional Disorders and Iowa Conners Rating Scale, were used as criterion measures. The following cutoff scores were identified using receiver operating characteristic curve (ROCC) analysis with good classification accuracies—Overall score: 15; Externalizing Problems and Attention Problems: 7; and Internalizing Problems: 5.

The PSC-17 as a Universal Screening Tool

The PSC-17 may be an optimal instrument to use for universal screening of young children. First, a universal screening scale should meet the needs of teachers and schools. Teachers should be able to complete the measure quickly, making the length a significant factor. Cost should be considered as well (Steed & Banerjee, 2016). Free of cost or low cost is an attractive feature for universal screening in schools. Ringwalt (2008) conducted a comprehensive review of screening scales of EBR for preschool children and noted that the PSC-17 is a brief and cost-efficient scale. In addition, the PSC-17 has similar items compared with other universal screening tools for preschool children (DiStefano, Liu, & Burgess, 2017). Despite these advantages (i.e., brief and free of cost), a high-quality instrument should integrate various validity evidence into a coherent account of the degree to which existing evidence and theory support the intended use of the scale (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014).

While the PSC-17 had been previously validated and successfully used in primary care settings with parent raters (e.g., Blucker et al., 2014; Chaffin et al., 2017; Erdogan & Ozturk, 2011; Gardner, Lucas, Kolko, & Campo, 2007; Murphy et al., 2016), additional research is needed to examine the psychometric properties of a measure before extending the use of the scale (AERA, APA, & NCME, 2014, Standard 1.4) to school-based settings. Until now, only one study has been conducted to examine the content validity and validate the factor structure of the PSC-17 for preschool children with teacher ratings (DiStefano et al., 2017). However, children are nested within classrooms with universal screening at school. The multilevel nature of the data has not been considered in analyzing the PSC-17 factor structure. In addition, no additional psychometric evidence was provided to support the usability of the PSC-17 in school-based settings for preschool children. The current study was targeted to fill this gap in the literature.

The purpose of the current study was to provide psychometric evidence of the PSC-17 for young children rated by teachers at school for universal screening (AERA, APA, & NCME, 2014). As correct identification of an instrument's structure is crucial to a scale's quality (Kline, 2014), we first examined the factor structure (i.e., construct validity) of the PSC-17. We intended to confirm the factor structure by including the factor structures of the child and the class level simultaneously. We also examined the cutoff score for the overall PSC-17 used to identify EBR. Finally, internal reliability and criterion validity were investigated to build additional psychometric support.

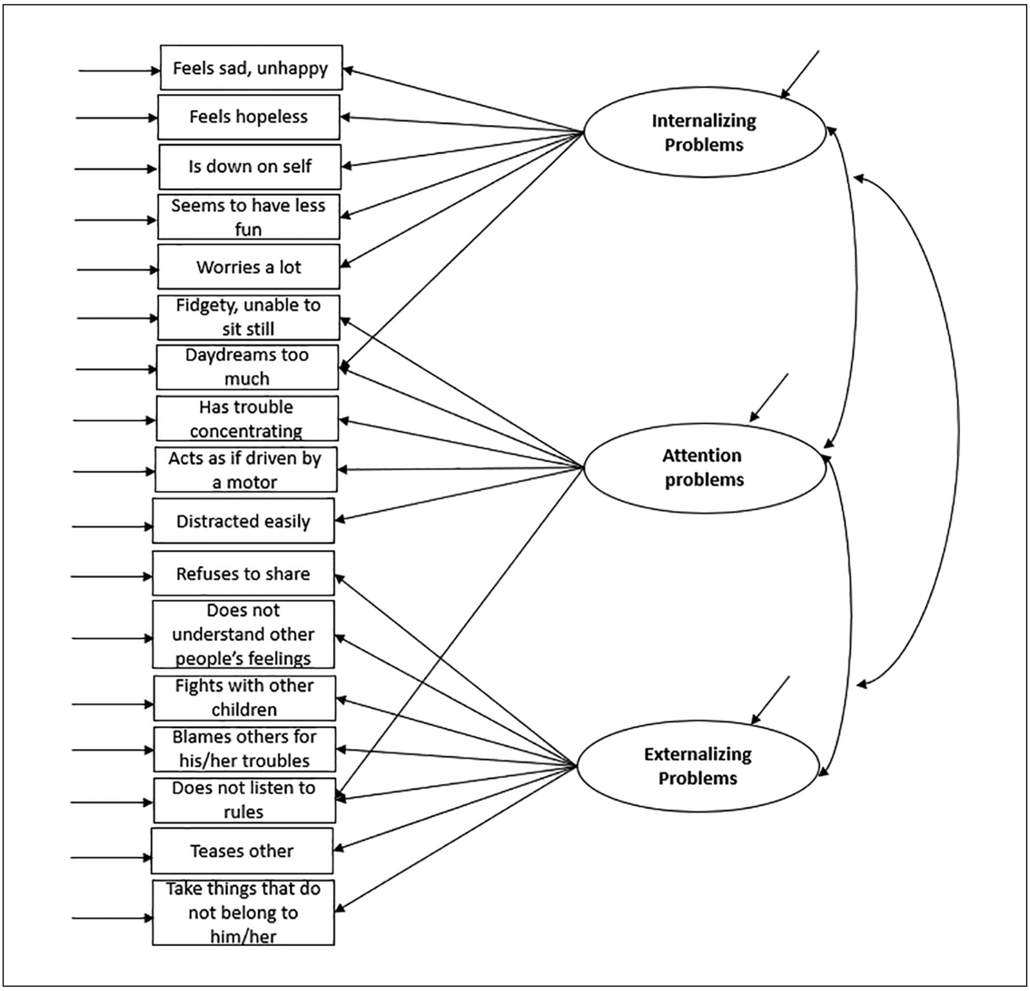

Figure 1. The factor structure of the PSC-17.
PSC-17 = Pediatric Symptom Checklist–17.

Method

Instrumentation

The PSC-17 was completed by 44 teachers in the Fall of 2012. Teachers provided ratings on all children in their classrooms for universal screening. Teachers rated occurrence of stated behaviors (see Figure 1) for each child. Items were rated on a 3-point ordinal scale with anchors of "never" = 0, "sometimes" = 1, and "often" = 2. Item scores were summed and higher scores indicate a higher level of EBR. EBR status was determined based on a cutoff score, and risk was noted if a child's score surpassed the cutoff score. We then randomly assigned teachers to two well-known comprehensive scales as an outcome measure of EBD—the BASC-2 (C. R. Reynolds & Kamphaus, 2004) or the ASEBA (Achenbach & Rescorla, 2001)—to obtain children's EBD information in the Spring of 2013. Random assignment of "long forms" was part of the study design to reduce the time burden on teachers. Both the ASEBA and the BASC-2 have national norms for score computation, and both forms are widely used by schools and psychologists nationwide as outcome measures in EBD identification (e.g., Chaffin et al., 2017). We considered two measures separately in cutoff score validation.

The BASC-2 Teacher Rating Scale–Preschool consists of 100 items (e.g., is nervous) for children aged 2 to 5 years. There were four response categories: *never, sometimes, often*, and *almost always* (scaled 0-3). The Behavioral Symptoms Index (BSI) was a measure of overall EBD level including information from the Hyperactivity, Aggression, Depression, Attention Problems, Atypicality, and Withdrawal scales. The test manual (C .R. Reynolds & Kamphaus, 2004) provided estimates of internal consistency, test–retest reliability and inter-rater reliability; all the estimates were above 0.70. Validity evidence was provided to assess the BASC-2's similarity to other similar behavioral scales, such as the ASEBA. The internal reliability of BSI in the current sample was .89 (N = 378).

The ASEBA Caregiver–Teacher Report Form is appropriate for children aged 1.5 to 5 years. Teachers or caregivers provide ratings on 99 items (e.g., hits others), regarding the frequency of behaviors. A 3-point ordinal subscale is used: *not true, somewhat true*, and *very true* (scaled 0-2). The overall scale measures six symptomatic subscales (Emotional Reactive, Anxious/Depressed, Somatic Complaints, Withdrawn, Attention Problems, and Aggressive Behaviors) and five *Diagnostic and Statistical Manual of Mental Disorders*–oriented scales (Affective Problems, Anxiety Problems, Pervasive Developmental Problems, Attention-Deficit/Hyperactivity Problems, and Oppositional Defiant Problems). The test manual (Achenbach & Rescorla, 2001) provided internal consistency estimates ranging from .66 to .96, inter-rater reliability ranging from .64 to .79, short-term test–retest reliability (i.e., 8-day test–retest interval) ranging from .77 to .88, and long-term test–retest reliability ranging from .40 to .64 (over a 3-month period). Validity evidence for the ASEBA was supported through content analysis in the selection of items, strong correlations with other similar scales, and ability to identify EBD status. The internal reliability of ASEBA total scores was .81 (N = 357) in the current sample.

Both the ASEBA and BASC-2 have reported t scores (M = 50, SD = 10). One standard deviation above the average t scores (i.e., significant t score $>$ 60) indicated the existence of EBD (Achenbach & Rescorla, 2001; C. R. Reynolds & Kamphaus, 2004). The results of the two forms are comparable with reported significant correlations (Achenbach & Rescorla, 2001; Greer, DiStefano, Liu, & Cain, 2015).

Participants

Forty-four teachers from eight elementary schools or child development centers (average poverty index: 80.14%) across six school districts in a southeastern state were involved in a federal grant study during the 2012-2013 academic year. Institutional Review Board approval and informed consent were obtained for each participating teacher before the study.[1] The schools were in suburban or rural areas, with an average class size of 19.57 children. Almost all of the 44 teachers were White females (N = 43, 97.72%); the average number of years of teaching experience was 6.75 years (SD = 6.46) at the early childhood level. Eighteen (40.91%) teachers held a bachelor's degrees, and 26 (59.09%) had achieved a master's degree.

Teachers' participation in the project was voluntary and teachers received a small stipend for participation. One issue in preschool screening is the inadequate training of teachers on form completion (Steed & Banerjee, 2016). During school visits, two brief training sessions were provided to the teachers who needed to complete scales to enhance data quality. All participating teachers attended the first training session in October of 2012. The research team reviewed the PSC-17, explained the meaning of each option, and answered questions. Teachers completed the PSC-17 by November of 2012. The second training was conducted in March 2013, and researchers reviewed the basics of the two outcome measures with the teachers. Teachers completed the BASC-2 or the ASEBA by April of 2013. The researchers revisited schools to collect completed packets. To limit missing data, researchers contacted teachers who failed to provide complete information in their responses. Finally, the researchers provided results sheets to schools or teachers.

A total of 738 young children were rated with the PSC-17. The children's average age was 4.70 years (*SD* = .60). Demographic information showed a roughly equal gender distribution of the rated children (female, *n* = 368, 49.86%; male, *n* = 370, 50.14%). The sample was multiracial, with predominantly White (*n* = 295, 39.97%), African American (*n* = 280, 37.94%), and Hispanic (*n* = 49, 6.64%) children represented. Most children received free or reduced lunch (*n* = 562, 76.15%). There was no missing data on the PSC-17; three children, without outcome data, were not considered in the cutoff score analysis.

Data Analysis

Multilevel confirmatory analysis and internal consistency. As children were nested within classes (i.e., teachers), multilevel analysis can be considered to distinguish the child level (i.e., within level) and the class level (i.e., between level). As the factor structure has been examined at preschool with teacher ratings (DiStefano et al., 2017), confirmatory factor analysis (CFA) was used to measure whether the same structure of the PSC-17 would be identified with a new sample, while taking into consideration the nesting of the data.

Multilevel confirmatory factor analysis (MCFA; Dedrick & Greenbaum, 2011) can be considered as an extension of the traditional CFA by investigating the factor structure of two levels simultaneously. In other words, different factor structures and item loading values can be obtained at the class and the child levels. MCFA decomposes a single covariance matrix in the traditional CFA into pooled class and child-level covariance matrices (Margola, Fenaroli, Sorgente, Lanz, & Costa, 2019). In the MCFA framework, this study investigates children's EBR at the individual and the class levels, respectively by accounting for both levels of variance simultaneously.

First, a CFA on the sample total covariance matrix (i.e., the traditional CFA) was conducted to provide guidance of potential MCFA models. Then, the portion of child-level variance relative to the total variance per item in both levels was examined with the intra-class correlations (ICCs). ICC values range from 0 to 1, with higher values indicating greater proportions of teacher-level variance in the sample. The multilevel nature of the data should be considered with high ICCs, typically values higher than .05 (Geldhof, Preacher, & Zyphur, 2014). The weighted least squares mean- and variance-adjusted estimator (WLSMV) was used within Mplus software (v. 7.4; Muthén & Muthén, 1998-2010) to accommodate the ordinal data obtained from the PSC-17.

Based on recommendations from other researchers, three MCFA models were considered (Dedrick & Greenbaum, 2011; Margola et al., 2019). First, a model of three factors at both levels, where factor loadings were freely estimated at both levels, was examined (Model 1). Next, the same model was tested by constraining the factor loadings to be equal at two levels (Model 2). Finally, we ran a model with the same factor structure at the child level and one single factor at the class level due to the small sample size at the class level (Model 3).

Six model fit indices were used to assess how well the data fit the proposed models: (a) ratio of chi-square value to degrees of freedom, (b) comparative fit index (CFI), (c) Tucker–Lewis Index (TLI), (d) root mean square error of approximation (RMSEA), and (e) weighted root mean square residual (WRMR). Chi-square values are often statistically significant with large sample sizes, thus the ratio of chi-square value to degrees of freedom can be used as a characteristic of model fit, with a ratio of less than 3 indicating an acceptable model fit (Schermelleh-Engel, Moosbrugger, & Müller, 2003). Both TLI and CFI are incremental fit indices and test the proportionate improvement in fit by comparing the target model to an independence model. TLI and CFI values of .95 or higher are indicative of excellent fit and values between .90 and .94 are typically seen as good fit (McDonald & Marsh, 1990). The RMSEA should approximate or be less than .08 to demonstrate close fit of the model (Hu & Bentler, 1999). Finally, the cutoff value of 1 is adequate for the WRMR (DiStefano, Liu, Jiang, & Shi, 2018). Cross-loading items and interpretability were considered based on the factor analysis structure identified from previous research (DiStefano et al., 2017). All proposed models were compared to find the best fitting model based on these indices. If ICCs were above the cutoff values, MCFA would be considered from a solid validity perspective to account for the nested feature of the data in spite of potential worse model fit. The standardized loadings of the optimal factor structure were provided, and correlations among subscales were reported. Finally, Cronbach's alpha was used to estimate the internal consistency of items in each factor for both the child and the class level. Values that were higher than .70 were acceptable (Frey, 2006).

Predictive validity and cutoff score analyses. The PSC-17 total score was used to predict the presence of EBD as determined by the BASC-2 or ASEBA scores. We validated the overall cutoff score instead of subscale cutoff scores, as schools are generally concerned about a child's overall EBR status with universal screening (Greer & Liu, 2016). A series of three analyses were conducted. First, basic descriptive information of the PSC-17 and outcome measures were reported by two subsamples (i.e., the PSC-17 with BASC-2 or ASEBA total scores). Next, predictive validity of the PSC-17 was examined by correlating the PSC-17 total scores and BASC-2 or ASEBA total scores. For validity estimates, values of 0.40 or higher were considered moderate and estimates of 0.80 or higher indicated a strong correlation (Hinkle, Wiersma, & Jurs, 2002).

Finally, ROCC procedures were used to examine viability of the PSC-17 recommended cutoff score in SPSS (v. 23.0). The same process was repeated twice using the BASC-2 and the ASEBA as the outcome measure separately in two subsamples. ROCC analysis is a popular method for creating optimal cut scores with psychological instruments (Swets, 1996), and it has been used in previous cutoff score validation of the PSC-17 (e.g., Gardner et al., 1999). This method attempts to minimize the number of classification errors (DiStefano & Morgan, 2011). Area under the curve (AUC) evaluates effectiveness of the PSC-17 as a screening tool by examining the accuracy of the total score to separate children with EBD from those without EBD. AUC values of 1 demonstrate a perfect test, AUC indices from 0.90 to 1 are considered excellent, and values between 0.80 and 0.90 are deemed good (Swets, 1996).

According to the outcome status (EBD or no EBD) and the screening status (EBR or no EBR), children can be placed into one of the four groups. The frequency of each group can then be calculated. True negative cases are children without EBR and identified by the screener as without EBD (a), while true positive cases are children with EBR correctly identified with EBD (d). False negative cases are children without EBR but exhibiting EBD later (b); false positive cases are children with EBR but classified without EBD (c). The combined number of all cases equals the total sample size.

Using frequency information, researchers typically calculate seven indices to investigate the accuracy of a cutoff score. Sensitivity and specificity are two commonly used indices. Sensitivity, or true positive rate, is the proportion of EBD children correctly identified from the screening results, $d / (b + d)$; specificity (true negative rate) is the probability of correctly identifying children who are without EBD, $a / (a + c)$. The false positive (FP) error rate is the proportion of children who are without EBD but are misclassified as cases with EBR, $c / (a + c)$. The false negative (FN) rate is the proportion of children who are with EBD but are misidentified as without EBR, $b / (b + d)$. The sum of FP and specificity is 1, and the sum of FN and sensitivity is 1. Finally, the positive predictive value (PPV) is the proportion of positive test results that are true positives, $d / (d + c)$. The negative predictive value (NPV) is the proportion of subjects with a negative test result who are correctly diagnosed, $a / (a + b)$. The total hit rate is the proportion of correctly identifying all negative and positive cases, $(a + d) / (a + b + c + d)$.

Table 1. PSC-17 MCFA Fit Indices.

Models	χ^2/df	RMSEA	CFI	TLI	WRMR
Model 1	763.49*/232 (3.29)	.06	.94	.92	1.32
Model 1 (cross-loadings)	665.30*/228 (2.92)	.05	.95	.94	1.18
Model 3	763.42*/235 (3.24)	.06	.94	.93	1.32
Model 3 (cross-loadings)	668.00*/233 (2.86)	.05	.95	.94	1.19

Note. PSC-17 = Pediatric Symptom Checklist–17; MCFA = multilevel confirmatory factor analysis; RMSEA = root mean squared error of approximation; CFI = comparative fit index; TLI = Tucker–Lewis fit index; WRMR = weighted root mean square residual.
*$p < .0001$.

The following guidelines (Glascoe, 2005) were used to evaluate the effectiveness of a cut score. The 0.80 criterion was used to balance true positive (sensitivity) and true negative (specificity) rates. We sacrificed specificity for sensitivity to minimize the FN rate because children with EBD who were missed at the early childhood level were not recoverable through later assessment. Conversely, children who were misclassified with EBR could be corrected through additional tests. Thus, sensitivity of 0.80 was used as the primary criterion. We considered recommending a cutoff score based on the minimum *d* value (Yovanoff & Squires, 2006), where $d = \sqrt{(1-\text{sensitivity})^2 + (1-\text{specificity})^2}$. Total hit rates of 0.80 or higher were thought to indicate good fit. Finally, there was no agreed standard for the positive predictive rate (PPV); thus, values were examined and interpreted in context of the study.

Results

Multilevel Confirmatory Analysis and Internal Consistency

The three-factor CFA model showed acceptable fit with most fit indices above the cutoff values, $\chi^2(116) = 638.43$, ratio = 5.89, RMSEA = .08, CFI = .96, TLI = .95, WRMR = 1.64. The CFA solution with two cross-loading items was identified as the optimal structure (Figure 1). This model showed good fit with all the indices above the cutoff values, $\chi^2(114) = 434.96$, ratio = 3.81, RMSEA = .06, CFI =.98, TLI =.97, WRMR = 1.30. Then, ICCs for all PSC-17 items ranged from .13 to .55, with 14 of the 17 items exhibiting values higher than .20.

Three MCFA models were considered based on the traditional CFA results by considering the cross-loading items. Although Model 2 (i.e., constrained factor loadings at both levels with the three-factor structure) yielded acceptable fit, it was not considered, as multiple Heywood cases were identified at the class level. Thus, fit indices in Model 1 (i.e., free estimation at both levels with the three-factor structure) and Model 3 (i.e., the three-factor structure at the child-level structure and one factor at the class level) were compared.

Both Model 1 and Model 3 with cross-loadings exhibited good fit to the data (Table 1). While all factor loadings were significant, class-level factor loadings were slightly higher under Model 1 with cross-loadings. After reviewing the available information, Model 1 with cross-loadings was selected as the optimal factor structure. The factor structure of Model 1 was provided in Figure 1, and both levels shared the same factor structure.

The MCFA standardized factor loadings were provided in Table 2. Almost all factor loadings were higher than .50, indicating moderate to strong relationships between items and factors (see Table 2). “Daydreams too much” loaded on Internalizing Problems and Attention Problems; “Does not listen to rules” loaded on Attention Problems and Externalizing Problems. Although the factor loadings were lower with cross-loading items, they were considered as acceptable based on the reasonable interpretability and improvement to model fit.

Table 2. PSC-17 Standardized Factor Loadings.

Items	Factors	Model 1 (with cross-loadings) Child	Model 1 (with cross-loadings) Teacher
Feels sad, unhappy	Internalizing Problems	.88	.92
Feels hopeless		.89	.88
Is down on self		.93	.98
Seems to have less fun		.75	.90
Worries a lot		.63	.91
Fidgety, unable to sit still	Attention Problems	.94	.95
Daydreams too much		.50 (.23)	.32 (.55)
Has trouble concentrating		.88	.96
Acts as if driven by a motor		.91	.80
Distracted easily		.92	.94
Refuses to share		.92	.75
Does not understand other people's feelings	Externalizing Problems	.89	.89
Fights with other children		.88	.81
Blames others for his or her troubles		.79	.99
Does not listen to rules		.57 (.55)	.62 (.20)
Teases other		.85	.88
Take things that do not belong to him or her		.85	.72

Note. Values in the parenthesis were factor loadings of two cross-loadings on the other factor. PSC-17 = Pediatric Symptom Checklist–17.

Table 3. Correlations of the PSC-17 Subscales.

	Internalizing and Attention Problems	Internalizing and Externalizing Problems	Attention and Externalizing Problems
Model 1 Child Level	.38	.52	.78
Model 1 Teacher Level	.84	.84	.82

Note. PSC-17 = Pediatric Symptom Checklist–17.

Correlations among factors illustrated strong relationships, except for the correlation between Internalizing and Attention Problems at the child level (see Table 3). In general, factor correlations were stronger at the class level than the child level. Finally, at the child level, the overall and subscales (Externalizing, Internalizing, and Attention Problems) yielded high internal consistency estimates (i.e., .91, .89, .81, and .88, respectively). At the class level, internal consistencies of overall and subscales (Externalizing, Internalizing, and Attention Problems) were .95, .91, .92, and .90, respectively.

Predictive Validity and Cutoff Scores Analysis

Descriptive information of the PSC-17 and outcome measures (BASC-2 or ASEBA scores) were provided in Table 4. Correlations between outcome scores and the PSC-17 total scores were of medium to high magnitude, suggesting that the PSC-17 total scores were associated with the outcome *t* scores (BASC-2: .66; ASEBA: .64).

The cutoff score for the PSC-17 total was validated using two measures separately (Table 5). Altogether, 82 (21.69%) children were identified with EBD using the BASC-2 BSI. Results from the ROCC analysis indicated that the PSC-17 could effectively distinguish between children with

Table 4. Descriptive Information of PSC-17 and Outcome Variable.

	Minimum	Maximum	*M*	*SD*
Subsample 1 (*N* = 378)				
PSC-17 Total	0	25.00	6.54	6.10
PSC-17 Internalizing Problems	0	10.00	1.50	2.11
PSC-17 Attention Problems	0	12.00	3.84	3.39
PSC-17 Externalizing Problems	0	12.00	2.24	2.88
BASC-2 BSI	36	103.00	51.74	11.40
Subsample 1 (*N* = 357)				
PSC-17 Total	0	32.00	4.62	5.83
PSC-17 Internalizing Problems	0	11.00	0.98	1.82
PSC-17 Attention Problems	0	12.00	2.87	3.27
PSC-17 Externalizing Problems	0	14.00	1.47	2.59
ASEBA Total	29	79.00	45.26	10.78

Note. Three children were deleted for cutoff score analysis due to missing outcome values. PSC-17 = Pediatric Symptom Checklist–17; BASC-2 = Behavior Assessment System for Children–2; BSI = Behavioral Symptoms Index; ASEBA = Achenbach System of Empirically Based Assessment.

Table 5. ROCC Analysis of the PSC-17 Cutoff Scores.

PSC cutoff scores	Sensitivity	Specificity	PPV	NPV	False positive rate	False negative rate	Overall hit rate
BASC-2: 7[a]	.79	.75	.46	.93	.25	.21	.76
BASC-2: 9[b]	.72	.83	.54	.91	.17	.28	.81
ASEBA: 5[a]	.88	.74	.26	.98	.26	.12	.75
ASEBA: 7[b]	.74	.82	.31	.97	.18	.26	.82

Note. ROCC = Receiver Operating Characteristic Curves; PSC-17 = Pediatric Symptom Checklist–17; PPV = positive predictive value; NPV = negative predictive value; BASC = Behavior Assessment System for Children; ASEBA = Achenbach System of Empirically Based Assessment.
[a]The cutoff scores were generated based on rules of minimum *d*, sensitivity of .8.
[b]The cutoff scores were generated based on rule of overall hit rate of .8.

and without EBD (AUC = .83). A cutoff score of 7 was suggested using the criterion of sensitivity of .80 and of minimum *d*; a cutoff score of 9 was suggested following the criterion of overall hit rate higher than .80. Thus, 140 (37.04%) and 109 (28.84%) children were identified with EBR with cutoff scores of 7 and 9, respectively. These cutoff scores were much lower than the recommended PSC-17 cutoff score of 15.

Next, the cutoff sore was validated using the same process with the ASEBA total score. Here, 9.52% of the children (*n* = 34) were identified with EBD suggested by the ASEBA results. ROCC results suggested exceptional fit as the AUC value was .86. The ROCC results suggested that a cutoff score of 5 following the sensitivity of .80 and minimum *d*, while a cutoff point of 7 was needed to obtain the overall hit rate of .80 with a sensitivity of .74. Again, the suggested cutoff scores were much lower than 15. Accordingly, 114 (31.93%) and 82 (22.97%) children were identified with EBR with cutoff scores of 5 and 7, respectively.

Discussion

The current study provided multiple important psychometric evidence to establish the usability of the PSC-17 in preschool. Our findings help extend usage of the PSC-17 with teacher ratings as a potential universal screening tool in the school environment.

Using MCFA, we identified a similar factor structure to that of previous studies (e.g., Blucker et al., 2014; DiStefano et al., 2017; Murphy et al., 2016). It was noted that the traditional CFA approach may present distorted results if the factor structures are different at two levels. The results supported construct validity of the PSC-17 and its underlying three-factor structure. The same factors (i.e., Externalizing Problems, Internalizing Problems, and Attention Problems) and cross-loadings (e.g., DiStefano et al., 2017) were confirmed at both the child and the class levels. In other words, the PSC-17 demonstrated invariance between children and teachers indicating that it can be used with nested data. However, the factor loadings and factor correlations were generally stronger at the class level than the child level. Although the findings support the three-factor structure of the PSC-17 at the class level, this remains a child-level construct as each student has their personal EBR status. If researchers are interested in the class-level factor structures for other research purposes, the scale may be considered as a class-level instrument. However, this is unusual in universal screening as the goal is to correctly identify EBR for each child.

High internal consistency estimates of the overall scale and subscales indicated stability of the PSC ratings at both levels. Predictive validity of the PSC-17 total was supported by moderate to high positive correlations between the PSC-17 total scores and BASC-2 or ASEBA total scores. The results also indicated viability of using the PSC-17 to identify EBR with a reasonable accuracy rate.

ROCC analyses suggested lower cutoff scores than those currently suggested with the PSC-17 total score (i.e., 15). It was noted that an overall cutoff score of 12 was identified in other PSC-17 studies for older children in primary care settings (Erdogan & Ozturk, 2011; Stoppelbein, Greening, Moll, Jordan, & Suozzi, 2012). The results appear reasonable as young children's observed behaviors may differ from older children in the original sample of the measure, and age differences should be considered in future cutoff score validation. The ASEBA subsample had fewer PSC-17 symptoms than the BASC-2 subsample, which might lead to slightly different cutoff scores between two subsamples. We suggest that a cutoff score of 7 for the PSC-17 total score may be a reasonable threshold for determining EBR with a screening instrument. The lower cutoff score was expected due to the sample difference between the original cutoff score creation in clinical settings and the current cutoff score validation in schools. Accurate cutoff scores with universal screening tools in the initial tier of RtI are necessary for EBD children being given comprehensive assessment and access to intervention services.

Finally, screening results can contribute to a collectively stable presentation of performance that is supportive of an EBD diagnosis; however, they cannot be used to diagnose EBD (Chaffin et al., 2017; Greer & Liu, 2016). The primary advantage of a universal screening scale is feasibility, brevity, and the ability to assess a large number of children efficiently. Although approximately 30% of children with a cutoff score of 7 were screened with positive EBR in the current study, the actual EBD prevalence rates were lower. The prevalence rates of EBR relate to the current sample, as no other studies have been conducted in the early childhood setting.

Limitations and Future Work

We are aware of the limitations of the present study. Initially, the current study used a sample from one state at one-time point and utilized a convenience sample. We hope that other researchers and school psychologists will consider replicating the study in other research settings to cross-validate the study results. We especially encourage researchers to validate the PSC-17 as a universal screening tool for older children, which may help building usability of the PSC-17 for longitudinal studies.

In addition, teachers rated children's behaviors in the current study. Although teachers are the optimal choice of universal screening in the school environment, teachers' ratings were the only data source used. As the PSC-17 was originally used with parents, future studies should also consider including both teachers and parents to address potential rater bias (Brauner & Stephens, 2006). Researchers have found that gender invariance was held in the PSC-17 and boys exhibited more problems than girls (Liu, DiStefano, Burges, & Wang, 2018). Differences in other demographic groups should be investigated in future studies. Finally, test scores and other outcomes (e.g., observations, classroom grades) may be used to support consequential validity of the PSC-17.

Two comprehensive outcome measures, BASC-2 BSI and ASEBA total scores, were treated as the "gold standard" for EBD identification. While these are well-accepted tools for this purpose (e.g., Chaffin et al., 2017), a true statement of EBD status may be preferred in future studies. However, this is extremely difficult for young children's EBD identification given the lack of a good standard (Brauner & Stephens, 2006). There was no good standard diagnostic tool available for EBD, which largely depends on the skills and collaboration of multiple stakeholders (Ogundele, 2018). Comprehensive measures may be the best choice in most circumstances. We only validated the overall cutoff score as subscale scores in outcome measures may not offer accurate EBD results. Children with EBR on the PSC-total may not have problems on all three subscales. Validation of subscale cutoff scores should be considered in future studies.

To expand usage of the PSC-17 to school-based settings, the current study used teacher ratings in preschool to examine its suitability of being a universal screening tool as part of the RtI system. Validating the PSC-17 in a new environment has made contributions for identifying EBD with young children. As the PSC-17 is a free and brief scale with sound psychometric evidence, the findings constitute an initial step toward providing evidence for the potential utility of the PSC-17 in the school setting. In addition, the researchers offered trainings to teachers to enhance the data quality. Finally, conducting universal screenings at school can help provide community support by offering services to children in need.

Acknowledgments

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R324A100104 to the South Carolina Research Foundation. The opinions expressed are those of the authors and do not necessarily represent views of the Institute of Education Sciences or the U.S. Department of Education.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Institute of Education Sciences – Grant #: R324A100104.

ORCID iD

Jin Liu https://orcid.org/0000-0002-4242-6582

Note

1. The study was conducted in compliance with ethical requirements of the cooperating agencies.

References

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms and profiles*. Burlington, VT: Research Center for Children, Youth, and Families.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Aos, S., Lieb, R., Mayfield, J., Miller, M., & Pennucci, A. (2004). *Benefits and costs of prevention and early intervention programs for youth*. Olympia: Washington State Institute for Public Policy.

Blucker, R. T., Jackson, D., Gillaspy, J. A., Hale, J., Wolraich, M., & Gillaspy, S. R. (2014). Pediatric behavioral health screening in primary care: A preliminary analysis of the Pediatric Symptom Checklist-17 with functional impairment items. *Clinical Pediatrics*, *53*, 449-455. doi:10.1177/0009922814527498

Brauner, C. B., & Stephens, C. B. (2006). Estimating the prevalence of early childhood serious emotional/behavioral disorders: Challenges and recommendations. *Public Health Reports*, *121*, 303-310. doi:10.1177/003335490612100314

Bullis, M., & Yovanoff, P. (2006). Idle hands: Community employment experiences of formerly incarcerated youth. *Journal of Emotional and Behavioral Disorders*, *14*(2), 71-85. doi:10.1177/10634266060140020401

Carta, J. J., & Greenwood, C. R. (2013). Future challenges and new directions in RTI. In V. Buysse & E. Peisner-Feinberg (Eds.), *Handbook of response to intervention in early childhood* (pp. 421-432). Baltimore, MD: Paul H. Brookes.

Chaffin, M., Campbell, C., Whitworth, D. N., Gillaspy, S. R., Bard, D., Bonner, B. L., & Wolraich, M. L. (2017). Accuracy of a pediatric behavioral health screener to detect untreated behavioral health problems in primary care settings. *Clinical Pediatrics*, *56*, 427-434. doi:10.1177/0009922816678412

Coleman, M. R., Roth, F. P., & West, T. (2009). *Roadmap to pre-k RTI: Applying response to intervention in preschool settings*. National Center for Learning Disabilities. Retrieved from https://www.ncld.org/wp-content/uploads/2014/11/roadmaptoprekrti.pdf

Dedrick, R. F., & Greenbaum, P. E. (2011). Multilevel confirmatory factor analysis of a scale measuring interagency collaboration of children's mental health agencies. *Journal of Emotional and Behavioral Disorders*, *19*(1), 27-40.doi:10.1177/1063426610365879

DiStefano, C., Liu, J., & Burgess, Y. (2017). Investigating the structure of the Pediatric Symptoms Checklist in the preschool setting. *Journal of Psychoeducational Assessment*, *35*, 494-505. doi:10.1177/0734282916647648

DiStefano, C., Liu, J., Jiang, N., & Shi, D. (2018). Examination of the weighted root mean square residual: Evidence for trustworthiness? *Structural Equation Modeling: A Multidisciplinary Journal*, *25*, 453-466. doi:10.1080/10705511.2017.1390394

DiStefano, C., & Morgan, G. (2011). Examining classification criteria: A comparison of three cut score methods. *Psychological Assessment*, *23*, 354-363. doi:10.1037/a0021745

Durbrow, E. H., Schaefer, B. A., & Jimerson, S. R. (2000). Learning behaviours, attention and anxiety in Caribbean children: Beyond the "usual suspects" in explaining academic performance. *School Psychology International*, *21*, 242-251. doi:10.1177/0143034300213002

Erdogan, S., & Ozturk, M. (2011). Psychometric evaluation of the Turkish version of the Pediatric Symptom Checklist-17 for detecting psychosocial problems in low-income children. *Journal of Clinical Nursing*, *20*, 2591-2599. doi:10.1111/j.1365-2702.2010.03537.x

Frey, B. (2006). *Statistics hacks: Tips & tools for measuring the world and beating the odds*. Sebastopol, CA: O'Reilly Media.

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 & Adolescent Psychiatry*, *46*, 611-618. doi:10.1097/chi.0b013e318032384b

Gardner, W., Murphy, J. M., Childs, G., Kelleher, K., Pagano, M., Jellinek, M., . . . Chiappetta, L. (1999). The PSC-17: A brief Pediatric Symptom Checklist with psychosocial problem subscales. A report from PROS and ASPN. *Ambulatory Child Health*, *5*, 225-236.

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*, 72-91. doi:10.1037/a0032138

Glascoe, F. P. (2005). Screening for developmental and behavioral problems. *Mental Retardation & Developmental Disabilities Research Reviews*, *11*, 173-179. doi:10.1002/mrdd.20068

Gokiert, R. J., Georgis, R., Tremblay, M., Krishnan, V., Vandenberghe, C., & Lee, C. (2014). Evaluating the adequacy of social-emotional measures in early childhood. *Journal of Psychoeducational Assessment*, *32*, 441-454. doi:10.1177/0734282913516718

Greenwood, C. R., Bradfield, T., Kaminski, R., Linas, M., Carta, J. J., & Nylander, D. (2011). The response to intervention (RTI) approach in early childhood. *Focus on Exceptional Children*, *43*(9), 1-22. doi:10.17161/fec.v43i9.6912

Greer, F. W., DiStefano, C. A., Liu, J., & Cain, L. K. (2015). Preliminary psychometric evidence of the Behavioral and Emotional Screening System Teacher Rating Scale–Preschool. *Assessment for Effective Intervention*, *40*, 240-246. doi:10.1177/1534508415571594

Greer, F. W., & Liu, J. (2016). Creating short forms and screening measures. In K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction: Standards and recent advancements* (pp. 272-287). Boston, MA: Hogrefe Publishers.

Gutman, L. M., Sameroff, A. J., & Cole, R. (2003). Academic growth curve trajectories from 1st grade to 12th grade: Effects of multiple social risk factors and preschool child factors. *Developmental Psychology*, *39*, 777-790. doi:10.1037/0012-1649.39.4.777

Hemmeter, M. L., Santos, R. M., & Ostrosky, M. M. (2008). Preparing early childhood educators to address young children's social-emotional development and challenging behavior: A survey of higher education programs in nine states. *Journal of Early Intervention*, *30*, 321-340. doi:10.1177/1053815108320900

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2002). *Applied statistics for the behavioral sciences* (5th ed.). Boston, MA: Houghton Mifflin Harcourt.

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

Ikeda, M. J., Neessen, E., & Witt, J. C. (2008). Best practices in universal screening. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (pp. 103-114). Bethesda, MD: National Association of School Psychologists.

Jones, D. E., Greenberg, M., & Crowley, M. (2015). Early social-emotional functioning and public health: The relationship between kindergarten social competence and future wellness. *American Journal of Public Health*, *105*, 2283-2290. doi:10.2105/AJPH.2015.302630

Kline, P. (2014). *An easy guide to factor analysis*. New York, NY: Routledge.

Lane, K. L., Little, A., Menzies, H., Lambert, W., & Wehby, J. (2010). A comparison of students with behavior challenges educated in suburban and rural settings: Academic, social, and behavioral outcomes. *Journal of Emotional and Behavioral Disorders*, *18*, 131-148. doi:10.1177/1063426609332343

Levitt, J. M., Saka, M., Romanelli, L. H., & Hoagwood, K. (2007). Early identification of mental health problems in schools: The status of instrumentation. *Journal of School Psychology*, *45*, 163-191. doi:10.1016/j.jsp.2006.11.005

Liu, J., DiStefano, C., Burgess, Y., & Wang, J. (2018). Pediatric Symptom Checklist-17: Testing measurement invariance of a higher-order factor model between boys and girls. *European Journal of Psychological Assessment*, 1-7. doi:10.1027/1015-5759/a000495

Margola, D., Fenaroli, V., Sorgente, A., Lanz, M., & Costa, G. (2019). The family relationships index (FRI): Multilevel confirmatory factor analysis in an Italian community sample. *European Journal of Psychological Assessment* , *35*, 335-345. https://doi.org/10.1027/1015-5759/a000427

Massachusetts General Hospital. (2007-2017). *Pediatric Symptom Checklist*. Retrieved from http://www.massgeneral.org/psychiatry/services/psc_use.aspx

McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin*, *107*, 247-255. doi:10.1037/0033-2909.107.2.247

McEvoy, A., & Welker, R. (2000). Antisocial behavior, academic failure, and school climate: A critical review. *Journal of Emotional and Behavioral Disorders*, *8*, 130-140. doi:10.1177/106342660000800301

Murphy, J. M., Bergmann, P., Chiang, C., Sturner, R., Howard, B., Abel, M. R., & Jellinek, M. (2016). The PSC-17: Subscale scores, reliability, and factor structure in a new national sample. *Pediatrics*, *138*(3), Article e20160038. doi:10.1542/peds.2016-0038

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Author.

Ogundele, M. O. (2018). Behavioral and emotional disorders in childhood: A brief overview for pediatricians. *World Journal of Clinical Pediatrics*, *7*(1), 9. doi:10.5409/wjcp.v7.i1.9

Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2001). Long-term effects of an early childhood intervention on educational achievement and juvenile arrest: A 15-year follow-up of low-income children in public schools. *Journal of American Medical Association*, *285*, 2339-2346. doi:10.1001/jama.285.18.2339

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior Assessment System for Children* (2nd ed.). Circle Pines, MN: American Guidance Service.

Ringwalt, S. (2008). *Developmental screening and assessment instruments with an emphasis on social and emotional development for young children ages birth through five*. National Early Childhood Technical Assistance Center, FPG Child Development Institute, The University of North Carolina at Chapel Hill. Retrieved from http://www.nectac.org/~pdfs/pubs/screening.pdf

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research-Online*, *8*(2), 23-74.

Steed, E. A., & Banerjee, R. (2016). Assessment and early identification of young children with social emotional difficulties and behavioral challenges. *Journal of Intellectual Disability-diagnosis and Treatment*, *3*, 198-204. doi:10.6000/2292-2598.2015.03.04.5

Stoppelbein, L., Greening, L., Moll, G., Jordan, S., & Suozzi, A. (2012). Factor analyses of the Pediatric Symptom Checklist-17 with African-American and Caucasian pediatric populations. *Journal of Pediatric Psychology*, *37*, 348-357. doi:10.1093/jpepsy/jsr103

Swets, J. A. (1996). *Signal detection theory and ROC analysis in psychology and diagnosis: Collected papers*. Mahwah, NJ: Lawrence Erlbaum.

Tandon, M., Cardeli, E., & Luby, J. (2009). Internalizing disorders in early childhood: A review of depressive and anxiety disorders. *Child and Adolescent Psychiatric Clinics of North America*, *18*, 593-610. doi:10.1016/j.chc.2009.03.004

Yovanoff, P., & Squires, J. (2006). Determining cutoff scores on a developmental screening measure: Use of receiver operating characteristics and item response theory. *Journal of Early Intervention*, *29*, 48-62. doi:10.1177/105381510602900104