Criterion validity of the Strengths and Difficulties Questionnaire (SDQ) with inpatient adolescents

Stephanie Kovacs [a], Carla Sharp [a,b,*]

[a] Department of Psychology, University of Houston, 126 Heyne Building, Houston, TX 77204-5502, United States
[b] Menninger Clinic, Houston, TX, United States

ABSTRACT

Increasing rates of adolescent admissions to inpatient psychiatric settings and acute length of stay necessitates valid psychiatric screening tools. The Strengths and Difficulties Questionnaire (SDQ) appears to have valuable clinical application due to its brevity and strong psychometric properties. In this study we aimed to evaluate the criterion validity of the SDQ in 159 psychiatric inpatients between the ages of 12–17 against the Computerized Diagnostic Interview Schedule for Children (CDISC). In determining the criterion validity of the SDQ against the CDISC-IV, we further sought to compare its classification accuracy with the DSM-oriented scales of two widely-used clinical screeners for children and adolescents, the Child Behavior Checklist (CBCL) and Youth Self Report (YSR). Results demonstrated comparable diagnostic accuracy for the three measures in detecting common emotional and behavioral disorders. Tentative clinical cutoffs were proposed specific to American adolescents for parent and youth reports. Sensitivity and specificity values are also reported and discussed.

© 2014 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Research has estimated that nearly 5% of American children have a diagnosable mental health condition (Simpson et al., 2005), and about half of all lifetime cases of mental illness begins by age 14 (Kessler et al., 2005). Inpatient facilities are often utilized for the treatment of adolescent psychopathology (Pottick et al., 2001; Connor et al., 2002), and rates of hospital admissions are rising (Blader, 2011). These admissions are usually time-limited as more stringent criteria for authorizing admissions and continued stay have led to a shift in the conceptualization of goals for inpatient hospitals from long-term treatment to acute crisis intervention and evaluation (Green and Jacobs, 1998; Blader, 2011). While length of admission stay has been reduced, frequency of lifetime admissions for adolescents has increased (Blader, 2011).

This shift has created a rise in patient turnover, leading to reduced clinical contact at inpatient facilities (Pottick et al., 2001). As a result, inpatient stays for youth are fast-paced and subsequently occur under less-than-ideal circumstances. For example, several adolescents in severe crisis may be admitted involuntarily at the same time. Often, admissions occur in the middle of the night when a limited number of clinical staff are present to oversee the process. These challenges pose serious barriers to the assessment process. Structured and standardized clinician-based interviews are considered the gold standard, but take extensive time to administer. They rely heavily on patient cooperation and trained clinical staff to administer, score, interpret, and integrate findings. Short, easy-to-administer questionnaire-based diagnostic screeners may serve a valuable role in formulating a tentative clinical picture, and while they can never replace a full clinical diagnostic interview, they may aid the treatment team while a more thorough assessment is underway.

The Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) has become internationally recognized as one of the most widely used screening instruments related to the mental health of children and adolescents (Vostanis, 2006). The SDQ consists of 25 items pertaining to social, emotional, and behavioral functioning of children and adolescents across five subscales: Conduct Problems, Inattention–Hyperactivity, Emotional Symptoms, Peer Problems, and Prosocial Behavior. It has been translated into more than 60 languages and is easily accessible online (http://www.sdqinfo.com). A recent review of 48 studies (Stone et al., 2010) has shown that the SDQ has promising reliability and validity across a variety of settings, both cross-sectionally and longitudinally (Sharp et al., 2005).

Unfortunately, nearly all of the psychometric work with the SDQ has been conducted with non-American samples. In our review, we

* Corresponding author at: Department of Psychology, University of Houston, 126 Heyne Building, Houston, TX 77204-5502, USA. Tel.: +1 713 743 8500; fax: +1 713 743 8633.
E-mail address: csharp2@central.uh.edu (C. Sharp).

could find only five studies in the psychological literature that utilized an American sample. None of these studies sampled a clinical population nor used the youth self-report (Dickey and Blumberg, 2004; Bourdon et al., 2005; Simpson et al., 2005; Palmieri and Smith, 2007; van Slyke, 2008). The lack of research with American and clinical samples is problematic because it limits the knowledge base for the SDQ to have wider applicability. Moreover, given the importance of assessing the validity across multiple settings and situations (Cronbach and Meehl, 1955), it is not enough to take validity research conducted with one population and assume the measure is equally applicable to another population under different circumstances, particularly with regard to cutoff scores. The lack of research on US samples also has practical implications. Currently, no American-based clinical cut-off scores are available for the self-report SDQ. Moreover, while parent-report (SDQ-P) cutoffs using an American community sample were being derived, service utilization was used as the criterion variable for validity testing (Bourdon et al., 2005; Simpson et al., 2005), which is not ideal given high levels of documented service underutilization of mental health (Szatmari et al., 1989; Zahner et al., 1992; Leaf et al., 1996; Verhulst and van der Ende, 1997; Simpson et al., 2005; Thurston and Phares, 2008). In summary, application of the SDQ in the United States could benefit from more focused research to establish a normative reference with American samples for both parent and youth versions of the SDQ.

Against this background, the current study had three aims. First, we aimed to evaluate the criterion validity of the SDQ in an American inpatient sample against the gold standard of an interview-based clinical diagnostic assessment tool, the Computerized Diagnostic Interview Schedule for Children-4th Edition (CDISC-IV; Shaffer et al., 2000). Second, in determining the criterion validity of the SDQ against the CDISC-IV, we further sought to compare SDQ's classification accuracy against two widely-used clinical screeners for children and adolescents, the Child Behavior Checklist (CBCL; Achenbach, 1991a) and Youth Self Report (YSR; Achenbach, 1991b; Achenbach and Rescorla, 2001). In so doing, we used the DSM-oriented scales of the CBCL/YSR rather than the syndrome scales, because the former have not yet been compared with the SDQ, while several European criterion validity studies of the SDQ have compared the SDQ with CBCL syndrome scales (Goodman and Scott, 1999; Klasen et al., 2000; Mullick and Goodman, 2001; Malmberg et al., 2003; Becker, Hagenberg et al., 2004; Becker, Woerner et al., 2004). Our third aim was to determine clinical cutoffs for parent and youth reports of the SDQ for use with American inpatient samples.

2. Methods

2.1. Participants

Participant recruitment was approved by the relevant Institutional Review Board. All participants and their parents were informed of the nature of the study and signed consent and assent forms in order to participate. Adolescents who chose not to participate in the study continued to receive treatment at the inpatient facility comparable to those who did participate.

A total of 174 consecutive submissions of adolescent patients and their parents were recruited from an inpatient psychiatric facility where the average length of stay is 4–6 weeks. Two patients revoked their consent during the assessment process and were removed from the study. The remaining 172 patients were examined for exclusion criteria. One patient was excluded due to lack of English fluency, two patients were removed due to active psychotic symptoms which disrupted the assessment, and eight patients were removed because they discharged prior to completion of the assessment materials. Two participants were removed because they were unable to complete enough CDISC items to yield a summary score. The final sample consisted of $n=159$ adolescents and their respective parent. Since the diagnostic groups on the CDISC were presented as separate modules, the sample total fluctuated substantially across analyses. The

reason for this fluctuation was that participants were not always consistent in their completion of the measures and specific modules of the CDISC.

Adolescents ranged in age from 12 to 17, with a mean age of 15.45 (S.D.=1.48). The sample consisted of 55% female and 45% male patients. Parent data consisted of the adolescent's legal guardian, most frequently a biological or adoptive parent. Ninety-three percent of the parents were mothers. Parents were asked to complete the interview independent of the other parent or family members to avoid bias. The ethnic distribution was 93.2% Caucasian, 1.2% African–American, 2.5% Hispanic, 1.2% Asian, and 0.6% Interracial.

The reason for hospital admission was determined by examining preliminary admission diagnoses using clinician ratings. Approximately 84% of the sample was admitted for mood disturbances, 42% had problems with anxiety, 0.04% displayed problems with psychosis, 27% had substance problems, 20% had difficulties with attention/hyperactivity, 0.06% displayed problems with opposition and defiance, and 0.07% were admitted for disordered eating.

Upon completion of the CDISC-P, the percent of diagnoses was distributed across the sample as follows: 40% anxiety disorders, 51% affective disorders (except Bipolar Disorder), 45% Oppositional Defiant Disorder (ODD), 20% Conduct Disorder (CD), and 30% Attention-Deficit/Hyperactivity Disorder (ADHD). Chi square analyses were conducted to examine rates of comorbidity in the sample. Of those who were diagnosed with an affective disorder, 27% also met criteria for an anxiety disorder ($\chi^2(1, N=146)=9.62, p < 0.01$), 28% also met criteria for Oppositional Defiant Disorder ($\chi^2(1, N=149)=7.46, p < 0.01$), 0.09% also met criteria for Conduct Disorder ($\chi^2(1, N=149)=0.97, p < 0.01$), and 14% also met criteria for ADHD ($\chi^2(1, N=149)=0.17, p=0.68$).

2.2. Measures

2.2.1. Strengths and Difficulties Questionnaire (SDQ)

The SDQ (Goodman, 1997) is a 25-item screening measure of emotional and behavioral disorders designed for children aged 3–17. Both parent and youth versions of the SDQ utilize a three-point Likert scale ranging from 0='not true', 1='somewhat true', to 2='certainly true'. The following five subscales can be derived from the measure and each comprised of five items: Conduct Problems, Inattention–Hyperactivity, Emotional Symptoms, Peer Problems, and Prosocial Behavior. A Total Difficulties score can also be derived by summing across the four problem behavior scales (i.e., Emotional Symptoms, Conduct Problems, Inattention–Hyperactivity, and Peer Problems). Items comprising the Emotional Symptoms scale include unhappy mood, fearfulness, headaches or stomachaches, clinginess, and worries. The Conduct Problems scale includes items pertaining to temper tantrums, obedience, lying or cheating, stealing, and fighting. The Inattention–Hyperactivity scale includes items pertaining to restlessness, fidgeting or squirming, distraction, concentration problems, impulsiveness, and task completion. The Peer Problems scale includes items regarding preference for solitary play, friendships, being liked or bullied, and preference for adults. The Prosocial scale includes consideration of others' feelings, sharing, displays of kindness, and willingness to help others.

The parent and youth reports of the SDQ were completed for the present study using the American English version for adolescents aged 11–17 (Bourdon et al., 2005). The SDQ has demonstrated acceptable reliability across a variety of settings, with weaker reliability for the Peer Problems and Prosocial Behavior scales. Research with American samples (parent report) has generated alphas ranging from 0.63 to 0.82 for the core problem scales, 0.46 to 0.62 for Peer Problems, and 0.71 to 0.77 for Prosocial Behavior (Bourdon et al., 2005; Palmieri and Smith, 2007).

2.2.2. Child Behavior Checklist (CBCL) and Youth Self Report (YSR)

The CBCL and YSR are a multiaxial empirically-based set of measures for assessing social competence and emotional/behavioral problems in children (Achenbach and Rescorla, 1991a, 1991b, 2001). The YSR was developed from the CBCL as a self-report measure for youths aged 11–18 (Achenbach and Rescorla, 1991b, 2001). Both measures are scored on a 3-point Likert scale (0='not true', 1='somewhat or sometimes true', and 2='very true or often true'). More recently, Achenbach et al. (2003) added DSM-oriented scales by reorganizing item-groupings. These scales were constructed via theoretical agreement with DSM-IV criteria and consisted of the following problem areas: Affective, Anxiety, Somatic, ADHD, ODD, and Conduct Problems. The DSM-oriented scales were used for the present study to compare their performance against the SDQ, which was also designed to reflect DSM-IV criteria. Several studies have reported positive psychometric properties of the DSM-oriented scales (Achenbach et al., 2003; Nakamura et al., 2009), while other research has found mixed results (van Lang et al., 2005; Vreugdenhil et al., 2006; Kendall et al., 2007). Convergent and divergent validity have been supported with several anxiety and depression scales (see Nakamura et al. (2009) for a summary of this validity research).

2.2.3. Computerized Diagnostic Interview Schedule for Children-4th Edition (CDISC-IV)

Both the parent and youth reports of the CDISC-IV (Shaffer et al., 2000) were used as the criterion measures due to their strict adherence to DSM-IV diagnostic criteria. The measure underwent development in 1979 by the National Institute of

Mental Health as a highly structured diagnostic instrument designed for use by nonclinicians and can be administered via paper–pencil or computer (Robins et al., 1981). The parent report was designed for parents/caregivers of children aged 6–17, and the youth report was designed for children aged 9–17. The interview is divided into a series of 24 modules pertaining to separate diagnoses that can be selectively administered based on the clinical or research question (see Shaffer et al. (2000) for a summary of the measure's development and reliability/validity findings). Most of the items can be answered with '*yes*', '*no*' and '*sometimes*' or '*somewhat*'. The interview is administered following computerized prompts that the interviewer reads out loud. The caregiver's answer is then inputted in the program and the program presents the next appropriate prompt. The interview therefore does not require clinical decision making on the part of the interviewer and inter-rater reliability is not calculated.

2.3. Procedures

Assessment questionnaires and interviews were typically administered within three days of admission but never beyond 2 weeks of hospital admission. Youth interviews were conducted face-to-face by a trained staff member. Parent interviews were conducted either via telephone by a trained graduate psychology student or face-to-face at the treatment facility. For consistency, all parent assessments were conducted with the same parent. All diagnoses were made within the first 2 weeks of each adolescent's admission.

The CDISC was administered by doctoral psychology students or clinical research assistants who had been trained and supervised by a licensed clinical psychologist. The interview was administered following computerized prompts read out loud by the interviewer. The child interviews were conducted on-site at the hospital, but most of the parent interviews were conducted via telephone, since many of the parents had flown their child in from out of state.

The CBCL and YSR were administered via computerized format under a licensed agreement. Since these measures are shorter than the CDISC, parents were able to complete their portion onsite at admission. Due to the computerized testing format, item-level data for the CBCL and YSR were unavailable and subsequently limited the ability to derive internal consistency data.

2.4. Data analytic strategy

To assess internal consistency (reliability) Cronbach's alpha was calculated. To evaluate criterion validity ROC analysis was used. With this method, a test measure is compared against a criterion to derive sensitivity and specificity values for determination of the measure's accuracy in detecting individuals with and without diagnoses. A two-dimensional ROC curve was graphed from scores of the test measure and criterion measure. The scores were then used to derive the sensitivity rate along the *Y* axis and the false positive rate (1-specificity) along the *X* axis. The area under the curve (*AUC*) is the most commonly used index of measuring accuracy of diagnostic classification and represents the ability of a test to correctly classify individuals with and without a given condition. An *AUC* of 1.0 represents ideal discrimination (100% sensitivity and 100% specificity), and an *AUC* of 0.50 represents chance where no discrimination exists. Graphically, the closer the ROC curve is to the upper left corner, the better the performance of the instrument. It is widely accepted that *AUCs* of 0.90–1.0 are considered excellent, 0.80–0.90 are good, 0.70–0.80 are fair/moderate, 0.60–0.70 are poor, and below 0.60 is no better than chance (Swets and Pickett, 1982). Rice and Harris (2005) compared *AUC* values with Cohen's *d* and determined that 0.71 corresponds to a *d* of 0.80, which is considered a large effect size by Cohen (1988). For the present study, a benchmark of 0.70 was set for *AUC* values as the minimum value necessary for meaningful detection ability. In order to compare *AUC* values, the DeLong et al. (1988) method was used.

Since the SDQ was created to reflect DSM criteria for several common disorders in childhood, the subscales were matched in the ROC analyses to detect the following CDISC diagnostic groups: the Emotional Symptoms scale matched to detect anxiety disorders (i.e., Social Phobia, Separation Anxiety Disorder, Specific Phobia, Panic Disorder, Agoraphobia, Generalized Anxiety Disorder, Obsessive Compulsive Disorder, and Post-Traumatic Stress Disorder) and affective disorders (i.e., Major Depressive Disorder and Dysthymic Disorder), the Conduct Problems scale to detect Oppositional-Defiant Disorder (ODD) and Conduct Disorder (CD), and the Inattention–Hyperactivity scaled matched to detect ADHD. Bipolar Disorder was not included in the affective disorders because the SDQ does not include items pertaining to manic symptoms. The Total Problems score was used to detect any of the preceding diagnostic groups. The presence of eating disorders, psychosis, substance abuse/dependence, learning disorders, developmental disabilities, etc. was not included in the determination of Axis I disorders, as the SDQ was not designed to detect these difficulties.

For consistency, the DSM scales of the CBCL and YSR were matched to the same diagnostic groups detailed for the SDQ. Unlike the SDQ, the CBCL and YSR have a separate scale for anxiety and affective disorders, and similarly ODD and CD. For this reason, separate analyses were conducted with the SDQ to determine its ability to detect each of these diagnostic groups independently for better comparison with the CBCL/YSR. The following scales of CBCL and YSR were used in the present study: Affective, Anxiety, ADHD, ODD, and Conduct Problems.

3. Results

3.1. Preliminary analyses

Descriptive statistics for the test measures are presented in Tables 1 and 2. The sample size across parent measures for ROC analysis ranged from 134 to 137 and 141 to 144 for the youth measures. Although the overall sample was much larger ($n = 159$), ROC analysis requires that the predictors and criterion measures must have complete data points across measures for each participant. If a participant completed only two of the three measures (e.g., CBCL and CDISC-P but not the SDQ-P), the procedure removed the participant for that particular analysis, even though there were partial data available. Since many of the inpatients had travelled to the hospital from out of state, there were inconsistencies in the ability to obtain data on every measure for every participant and his or her parent. Unlike ROC analysis, descriptive statistics on each scale can be completed independent of available data on other measures. As such, the sample sizes in Table 2 are much larger than in Table 1. Nonetheless, there were minor challenges in data collection which contributed to the noted fluctuations in sample size across predictor measures (e.g., noncooperation, fatigue, etc.).

Overall, Cronbach alphas for the SDQ were acceptable based on the recommended rubric by Cicchetti (1994) for interpreting reliability coefficients. Of the 12 scales examined, the Emotional Symptoms (parent report) and Peer Problems (parent and youth reports) displayed alphas below 0.70. An existing literature on the Peer Problems scale reflects consistently low alphas, most likely due to problems with the factor structure (Goodman, 1997, 2001; Goodman et al., 1998; Hawes and Dadds, 2004; Mellor, 2004; Bourdon et al., 2005; Becker et al., 2006; Palmieri and Smith, 2007).

Intercorrelations between the scales of the criterion and predictor variables are presented in Table 2. Correlations between scales of similar problem areas of the CBCL and SDQ-P ranged from low to moderate. For comparisons between the YSR and SDQ-Y, correlations were in the low to high range. Within the SDQ-P and SDQ-Y, correlations between scales were generally in the low to moderate range as well. As would be expected, the higher the correlation between scales on the predictor and criterion variables, the more similar their *AUC* values in the ROC analysis.

3.2. Criterion validity analysis

ROC analyses showed that in general, *AUC* values for the SDQ-P/Y and CBCL/YSR were in the fair to good range of diagnostic detection. Only the *AUC* value for Emotional Symptoms of the SDQ-P detecting affective disorders did not meet the benchmark of 0.70 ($AUC = 0.68$). The SDQ-P and Y generally detected difficulties as well as the CBCL and YSR, with the exception of parent-reported ADHD symptoms and youth-reported conduct problems. Results are displayed in Table 3.

3.2.1. Proposed cutoffs

ROC analysis was used to propose tentative cutoffs for use with clinical samples in the United States. Clinical and borderline cutoffs for the SDQ-P and Y are presented in Table 4, with corresponding sensitivity, specificity, Positive Predictive Values (PPV), and Negative Predictive Values (NPV) for the borderline and clinical ranges. Clinical cut points reflect the intersection of sensitivity and specificity values, which indicates the test score that yielded greatest accuracy of detecting individuals with and without a diagnosis. Borderline values or ranges are also reported and are likely to consist of subclinical symptoms. It is important to remember that although inpatient samples will

yield higher scores on measures of emotional and behavioral problems compared to community samples, clinical severity across all domains of difficulty should not be assumed.

Therefore, a borderline range of subthreshold difficulties still has merit in an inpatient population.

Only cutoffs relating to scales in which the *AUC* benchmark was achieved are displayed. Clinical cutoffs were determined by plotting sensitivity and specificity values graphically and matching their intersection with a subtest score, which reflected optimization of both sensitivity and specificity. In general, sensitivity and specificity values for clinical cutoffs were in the mid 60s and 70s, with a few values in the 80s. PPV and NPV spanned a larger range from the high 30s to the low 90s.

4. Discussion

The overarching aim of the current study was to examine the criterion validity of the SDQ in an American sample of inpatient adolescents. In all, the SDQ performed adequately, but our study also revealed several areas for improvement in terms of reliability as well as criterion validity. Regarding reliability, coefficient alpha estimates of the SDQ were generally acceptable against the background of Reynolds et al. (2008) observations that there are multiple factors for considering a reliability coefficient such as the construct being measured, time available for testing, how the scores are to be used, amongst others, so that reliability coefficients as low as 0.60 are acceptable. Therefore, in general the reliability estimates reported here were within acceptable range. The Peer Problems scale, however, displayed a considerably low alpha for both the parent and youth reports. This scale has generated notorious difficulties in the literature with internal consistency (Goodman, 1997, 2001; Goodman et al., 1998; Hawes and Dadds, 2004; Mellor, 2004; Bourdon et al., 2005; Becker et al., 2006; Palmieri and Smith, 2007). In their evaluation of the factor structure of the SDQ, Palmieri and Smith suggested that the low

Table 1
Summary statistics for SDQ-(parent/youth) and CBCL/YSR.

Subscale	Mean	Std. deviation	α	Scale range
SDQ-P ($n=152$)				
Total Difficulties	18.44	6.19	0.76	0–40
Emotional Symptoms	5.58	2.55	0.68	0–10
Conduct Problems	3.64	2.54	0.74	0–10
Inattention–Hyperactivity	6.06	2.41	0.72	0–10
Peer Problems	3.16	2.07	0.54	0–10
Prosocial Behavior	6.76	2.31	0.75	0–10
SDQ-Y ($n=159$)				
Total Difficulties	16.66	7.37	0.85	0–40
Emotional Symptoms	5.03	2.92	0.79	0–10
Conduct Problems	3.12	2.40	0.71	0–10
Inattention–Hyperactivity	5.77	2.78	0.80	0–10
Peer Problems	2.75	2.07	0.57	0–10
Prosocial Behavior	8.08	1.87	0.70	0–10
CBCL ($n=149$)				
Total Problems	73.62	26.65		0–110
Affective Problems	11.64	4.86		0–26
Anxiety Problems	5.03	2.96		0–12
ODD Problems	5.27	2.73		0–10
Conduct Problems	8.12	5.90		0–34
ADHD Problems	6.93	3.38		0–14
YSR ($n=156$)				
Total Problems	69.27	31.16		0–106
Affective Problems	10.91	6.40		0–26
Anxiety Problems	4.85	3.14		0–12
ODD Problems	4.87	2.46		0–10
Conduct Problems	7.19	5.16		0–30
ADHD Problems	7.08	3.29		0–14

Table 2
Pearson correlations between SDQ-P/Y and CBCL/YSR scales.

Measure/subscale	SDQ subscales					
	Total Difficulties	Emotional Symptoms	Conduct Problems	Inattention–Hyperactivity	Peer Problems	Prosocial Behavior
CBCL						
Total Problems	**0.71****	0.45**	0.56**	0.53**	0.29**	−0.27**
Affective Problems	0.31**	**0.55****	0.01	0.06	0.18*	−0.11
Anxiety Problems	0.41**	**0.64****	0.01	0.21*	0.22**	0.11
ODD Problems	0.51**	0.04	**0.76****	0.43**	0.08	**−0.40****
CD Problems	0.44**	−0.09	0.77**	0.42**	0	**−0.43****
ADHD	0.69**	0.20*	0.56**	**0.78****	0.23**	−0.19*
YSR						
Total Problems	**0.83****	0.72**	0.56**	0.61**	0.49**	−0.18*
Affective Problems	0.67**	**0.74****	0.25**	0.44**	0.48**	−0.13
Anxiety Problems	0.62**	**0.75****	0.18*	0.34**	0.51**	0.01
ODD Problems	0.47**	0.23**	**0.63****	0.36**	0.15	**−0.21****
CD Problems	0.42**	0.10	**0.65****	0.38**	0.12	**−0.35****
ADHD	0.71**	0.43**	0.61**	**0.76****	0.22**	−0.12
SDQ-P						
Total Difficulties		0.59**	0.68**	0.77**	0.54**	−0.36**
Emotional Symptoms			0.04	0.24**	0.21**	0.06
Conduct Problems				0.55**	0.12	−0.54**
Inattention–Hyperactivity					0.17*	−0.21*
Peer Problems						−0.24**
SDQ-Y						
Total Difficulties		0.80**	0.68**	0.79**	0.60**	−0.21**
Emotional Symptoms			0.26**	0.51**	0.45**	0
Conduct Problems				0.50**	0.22**	−0.39**
Inattention–Hyperactivity					0.19**	−0.05
Peer Problems						−0.21**

Bolded values reflect scales of similar problem areas.
Note: All correlations reflect comparisons of consistent informant type (e.g., CBCL with SDQ-P and YSR with SDQ-Y).

* $p \leq 0.05$.
** $p \leq 0.01$.

internal consistency of the Peer Problems scale may be due to the fact that two reverse-scored items comprise the scale, which could contribute to measurement error. In our study, the low measure of internal consistency for the Peer Problems scale was evident for both the parent and youth reports. Consequently, we recommend against the use of this scale, especially since it was not designed to measure psychopathology per se.

Only one of the psychopathology scales yielded an alpha below 0.70, the Emotional Symptoms scale for the parent report. Research has shown that when parents report internalizing disorders for their children, they are generally less consistent than youth-reported information for such disorders (Grills and Ollendick, 2002). Parents tend to be better reporters of externalizing disorders, most likely because the symptoms are easier to observe behaviorally. The low coefficient alpha estimate for the Emotional Symptoms scale was only for the parent report, which suggests that the parents in our sample were less adept at evaluating their child's internalizing symptoms compared to their externalizing problems. Low coefficient alpha estimates should also be interpreted against the background of new and better approaches to determine reliability through the use of measurement models.

In terms of criterion validity, results of the ROC analyses suggested that all AUC values were above 0.70, but that more than half of the reported AUC values were below 0.80. While Rice and Harris (2005) determined that an AUC value of 0.71 corresponds to d of 0.80, which is considered a large effect size by Cohen (1988), the consensus is that AUCs between 0.70 and 0.80 should be considered fair/moderate. The question then arises as to what additional psychometric work should be conducted with the SDQ in American samples to further establish its validity. The AERA, APA, NCME Standards for Educational and Psychological Testing suggested a scheme for organizing sources of evidence to evaluate proposed interpretations of test scores, including evidence based on test content, response processes, internal structure, relations to other (external) variables and consequences of testing (American Psychiatric Association, 1999). The current study focused on evidence in support of the SDQ based on relations to other (external) variables in the form of interview-based diagnosis. All other aspects of validity are worthy of further study, especially the internal factor structure of the SDQ, as studies using factor analytic approaches in clinical American samples are seriously lacking in this regard. Validity evidence based on internal structure (i.e., the degree to which the relationships among the component parts of the SDQ conform to the hypothesized constructs) is crucial to further establish its use in clinical samples in the United States. In addition, cross-cultural study methodology may be utilized to examine the cultural validity of the SDQ. Given similarities in culture between the United States and the United Kingdom (where

Table 3
Results of ROC analysis for criterion validity of SDQ (parent/youth) and CBCL/YSR.

Criterion	Scale	AUC	Scale	AUC	p	n
CDISC-P	SDQ-P		CBCL			
Any Axis I	Total Difficulties	0.77	Total Problems	0.75	0.69	136
Anxiety	Emotional Symptoms	0.77	Anxiety	0.75	0.65	134
Affective	Emotional Symptoms	0.68	Affective	0.71	0.45	137
ODD	Conduct Problems	0.80	ODD	0.85	0.12	137
CD	Conduct Problems	0.88	CD	0.88	0.92	137
ADHD	Inattention–Hyperactivity	0.74	ADHD	0.81	0.05	137
CDISC-Y	SDQ-Y		YSR			
Any Axis I	Total Difficulties	0.86	Total Problems	0.87	0.89	144
Anxiety	Emotional Symptoms	0.75	Anxiety	0.79	0.20	142
Affective	Emotional Symptoms	0.79	Affective	0.82	0.35	141
ODD	Conduct Problems	0.79	ODD	0.80	0.92	141
CD	Conduct Problems	0.80	CD	0.90	0.01	141
ADHD	Inattention–Hyperactivity	0.76	ADHD	0.77	0.90	141

Note: p-values reflect significance test between area under the curve (AUC) values of SDQ-P/Y and CBCL/YSR.

Table 4
Tentative cutoffs for SDQ-P and Y with clinical samples in the U.S.A.

	Proposed cutoffs			Sensitivity/specificity and PPV/NPV (borderline)				Sensitivity/specificity and PPV/NPV (clinical)			
	Normal	Borderline	Clinical	Se	Sp	PPV	NPV	Se	Sp	PPV	NPV
SDQ-P											
Total Difficulties	0–12	13–14	15–40	81.48	42.31	85.28	35.77	73.15	73.08	91.77	39.89
Emotional Symptoms (Anx)	0–3	4	5–10	75.36	51.47	51.28	75.50	60.87	61.76	51.90	69.96
Conduct Problems (ODD)	0–1	2	3–10	84.37	66.25	84.37	66.25	73.44	73.75	68.99	77.74
Conduct Problems (CD)	0–2	3	4–10	89.66	63.48	37.29	96.20	79.31	75.65	44.10	93.79
Inattention–Hyperactivity	0–3	4–5	6–10	95.35	32.98	38.10	94.25	67.44	64.89	45.39	82.16
Prosocial Behavior (ODD)	9–10	7–8	0–6	73.44	55.00	56.48	72.25	59.38	70.00	61.15	68.42
Prosocial Behavior (CD)	8–10	5–7	0–4	55.17	79.13	39.04	87.93	48.28	89.57	52.86	87.73
SDQ-Y											
Total Difficulties	0–11	12	13–40	80.19	66.67	88.96	50.13	76.42	75.76	91.35	48.97
Emotional Symptoms (Anx)	0–2	3	4–10	82.05	50.77	67.25	69.66	74.36	64.62	72.14	67.16
Emotional Symptoms (Aff)	0–2	3	4–10	92.06	51.90	59.87	89.35	82.54	63.29	63.67	82.30
Conduct Problems (ODD)	0–1	2–3	4–10	84.85	57.66	37.31	92.76	66.67	83.78	54.97	89.43
Conduct Problems (CD)	0–1	2	3–10	84.85	57.66	37.31	92.76	75.76	71.17	43.84	90.81
Inattention–Hyperactivity	0–4	5–6	7–10	89.29	53.51	32.58	95.21	64.29	73.68	38.06	89.13
Peer Problems (Anx)	0–1		2–10	78.75	52.24	67.01	66.61	63.75	68.66	71.48	60.59

Note: Abbreviations in the table are as follows: Se (sensitivity), Sp (specificity), PPV (positive predictive value), and NPV (negative predictive value).

the SDQ was developed) one would assume cultural invariance, but this is an empirical question that requires future attention.

Beyond findings regarding the overall reliability and criterion validity discussed above, two other findings are of note. First, the SDQ-Y was more accurate than the SDQ-P at identifying psychopathology, which is not surprising given well-established research that suggest that adolescents may be better reporters of their own psychopathology than their parent (Grills and Ollendick, 2002). Second, regardless of the reporter, the SDQ functioned remarkably similar to the CBCL/YSR at classifying psychopathology – that is, the CBCL and YSR yielded relatively low *AUC* values as well. One possibility for this finding is the severity and comorbidity of the inpatient sample in the current study. While it is important to evaluate the validity and reliability of assessment tools in real-life samples, it is also true that groups derived from the CDISC were likely not pure in psychopathology, resulting in lower *AUC* values for the SDQ, CBCL and YSR. While all measures performed relatively similarly, there were two exceptions where the CBCL/YSR outperformed the SDQ, namely with the Inattention–Hyperactivity scale of the SDQ-P and the Conduct Problems scale of the SDQ-Y. Taken together, we recommend using the self-report version of the SDQ with adolescents, excluding the Peer Problems subscale. If parent data can be collected on conduct problems, this would be ideal to use in conjunction with the SDQ-Y, since the SDQ-P was as accurate as the CBCL in classifying Conduct Disorder. Although the SDQ-Y still yielded a high *AUC* value of 0.80 with the Conduct Problems subscale, it was not as accurate as the YSR in this area.

Results of the present study should be considered a first-step in expanding knowledge of the SDQ as a screener in American inpatient units. While results are promising, they must be replicated with larger samples and with a wider range of ethnic distribution. Limitations of the present study are that the data consisted primarily of Caucasian adolescents and were primarily upper middle-class. Consequently, results may not be generalizable to other inpatient samples. Related, given that parent report constituted part of the criterion ratings in the current study, it would have been valuable to obtain to demographic information on parents. An additional limitation was the high rates of comorbidity among individuals diagnosed with an affective disorder. While comorbidity reflects the reality of inpatient samples, it also complicates the efficiency of the ROC analysis. In addition, future validity testing of the SDQ should continue to use gold-standard diagnostic measures as the criterion variable and incorporate teacher-report along with younger aged youth. Furthermore, we acknowledge that PPV/NPV and sensitivity/specificity values reported were somewhat low for the SDQ. The cutoffs proposed in Table 4 were derived from an effort to balance sensitivity and specificity, with slight favor towards sensitivity. Screening measures typically favor sensitivity in order to better ensure that problem areas are not overlooked. Best practice would be to follow-up with additional assessments, preferably those with higher specificity in order to rule out any false positives identified in the screener. In the present study, cutoffs were established the same way regardless of whether the report came from the parent or youth. Since parents tend to be better identifiers of externalizing problems, and youth tend to be better reporters of internalizing problems (Grills and Ollendick, 2002), we considered parent and youth reports as equally essential and unique in contributing to the initial screening process. Therefore, we maintained consistency in the preference we gave to sensitivity vs. specificity for the proposed cutoffs between informants. We highlight that these cutoffs are tentative and require replication studies with larger and more diverse samples.

Notwithstanding these limitations, this is the first study in any country to examine the 'DSM-based' scales of the SDQ and CBCL/

YSR against an interview-based diagnostic tool. Given that the DSM-5 (American Psychiatric Association, 2013) made no substantive changes to the symptom clusters evaluated here, it implies that the SDQ is still relevant even with a revised diagnostic manual. This is important because clinicians are more likely to use DSM-based scales, as they are conceptually more in line with the DSM itself. This is also the first study in any country to examine the relationship between subscales of the SDQ and CBCL/YSR against specific DISC-generated diagnoses. Other studies have compared the SDQ-P and Y with the CBCL/YSR by grouping diagnoses together broadly or by detecting only the presence/absence of psychopathology rather than specific diagnostic domains. We believe that the findings generated by this study offer preliminary support for use of the SDQ in inpatient hospitals with American youth, pending further validation work on the internal structure of the SDQ and its cross-cultural validity. Pending such work, the SDQ may be the first step in a diagnostic procedure whereby a nurse administers the SDQ. This information is then consulted during the in-take psychiatric interview by the attending physician who can confirm, disconfirm and elaborate information gained through the SDQ. Where an adolescent lies in terms of population-based norms can be discussed and integrated into the clinical formulation during team diagnostic conference. Finally, the SDQ can be administered at discharge to assess whether any change took place during the hospital stay (or follow-up if referred to a step-down program).

Acknowledgment

References

Achenbach, T.M., 1991a. Manual for the Child Behavior Checklist: 4-18 and 1991 Profile. Department of Psychiatry, University of Vermont, Burlington, VT.

Achenbach, T.M., 1991b. Manual for the Youth Self-Report and 1991 Profile. Department of Psychiatry, University of Vermont, Burlington, VT.

Achenbach, T.M., Dumenci, L., Rescorla, L.A., 2003. DSM-oriented and empirically based approaches to constructing scales from the same item pools. Journal of Clinical Child and Adolescent Psychology 32 (3), 328–340.

Achenbach, T.M., Rescorla, L.A., 2001. Manual for the ASEBA School-Age Forms and Profiles. Research Center for Children, Youth, and Families, University of Vermont, Burlington, VT.

American Psychiatric Association, 1999. Standards for Educational and Psychological Testing. American Educational Research Association, Washington, DC.

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders, 5th ed. American Psychiatric Publishing, Arlington, VA.

Becker, A., Hagenberg, N., Roessner, V., Woerner, W., Rothenberger, A., 2004. Evaluation of the self-reported SDQ in a clinical setting: do self-reports tell us more than ratings by adult informants? European Child and Adolescent Psychiatry 13 (Suppl 2), SII17–SII24.

Becker, A., Steinhausen, H.C., Baldursson, G., Dalsgaard, S., Lorenzo, M.J., Ralston, S.J., Dopfner, M., Rothenberger, A., 2006. ADORE Study Group, 2006. Psychopathological screening of children with ADHD: Strengths and Difficulties Questionnaire in a pan-European study. European Child and Adolescent Psychiatry 15 (Suppl 1), S56–S62.

Becker, A., Woerner, W., Hasselhorn, M., Banaschewski, T., Rothenberger, A., 2004. Validation of the parent and teacher SDQ in a clinical sample. European Child and Adolescent Psychiatry 13 (Suppl 2), SII11–SII16.

Blader, J.C., 2011. Acute inpatient care for psychiatric disorders in the United States, 1996 through 2007. Archives of General Psychiatry 68 (12), 1276–1283.

Bourdon, Goodman, K.H., Rae, R., Simpson, D.S., Koretz, D.S., G., 2005. The Strengths and Difficulties Questionnaire: U.S. normative data and psychometric properties. Journal of the American Academy of Child and Adolescent Psychiatry 44 (6), 557–564.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Routledge Academic, Hillsdale.

Cicchetti, D.V., 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment 6 (4), 284–290.

Connor, D.F., Miller, K.P., Cunningham, J.A., Melloni Jr., R.H., 2002. What does getting better mean? Child improvement and measure of outcome in residential treatment. The American Journal of Orthopsychiatry 72 (1), 110–117.

Cronbach, L., Meehl, R., 1955. Construct validity in psychological tests. Psychological Bulletin 52 (4), 281–302.

DeLong, E.R., DeLong, D.M., Clarke-Pearson, D.L., 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44, 837–845.

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 and Adolescent Psychiatry 43 (9), 1159–1167.

Goodman, R., 1997. The Strengths and Difficulties Questionnaire: a research note. Journal of Child Psychology and Psychiatry 38 (5), 581–586.

Goodman, R., 2001. Psychometric properties of the Strengths and Difficulties Questionnaire. Journal of the American Academy of Child and Adolescent Psychiatry 40 (11), 1337–1345.

Goodman, R., Meltzer, H., Bailey, V., 1998. The Strengths and Difficulties Questionnaire: a pilot study on the validity of the self-report version. European Child and Adolescent Psychiatry 7, 125–130.

Goodman, R., Scott, S., 1999. Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: is small beautiful? Journal of Abnormal Child Psychology 27 (1), 17–24.

Green, J., Jacobs, B., 1998. Current challenges. In: Green, J., Jacobs, B. (Eds.), Inpatient Child Psychiatry: Modern Practice, Research and the Future. Routledge Academic, London, pp. 3–8.

Grills, A.E., Ollendick, T.H., 2002. Issues in parent-child agreement: the case of structured diagnostic interviews. Clinical Child and Family Psychology Review 5 (1), 5–83.

Hawes, D.J., Dadds, M.R., 2004. Australian data and psychometric properties of the Strengths and Difficulties Questionnaire. Australian and New Zealand Journal of Psychiatry 38 (8), 644–651.

Kendall, P.C., Puliafico, A.C., Barmish, A.J., Choudhury, M.S., Henin, A., Treadwell, K. S., 2007. Assessing anxiety with the Child Behavior Checklist and the Teacher Report Form. Journal of Anxiety Disorders 21 (8), 1004–1015.

Kessler, R.C., Chiu, W.T., Demler, O., Walters, E.E., 2005. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Archives of General Psychiatry 62 (6), 617–627.

Klasen, H., Woerner, W., Wolke, D., Meyer, R., Overmeyer, S., Kaschnitz, W., Rothenberger, A., Goodman, R., 2000. Comparing the German versions of the Strengths and Difficulties Questionnaire (SDQ-Deu) and the Child Behavior Checklist. European Child and Adolescent Psychiatry 9 (4), 271–276.

Leaf, P.J., Alegria, M., Cohen, P., Goodman, S.H., 1996. Mental health service use in the community and schools: results from the four-community MECA study. Journal of the American Academy of Child and Adolescent Psychiatry 35 (7), 889–897.

Malmberg, M., Rydell, A.M., Smedje, H., 2003. Validity of the Swedish version of the Strengths and Difficulties Questionnaire (SDQ-Swe). Nordic Journal of Psychiatry 57 (5), 357–363.

Mellor, D., 2004. Furthering the use of the Strengths and Difficulties Questionnaire: reliability with younger child respondents. Psychological Assessment 16 (4), 396–401.

Mullick, M.S.I., Goodman, R., 2001. Questionnaire screening for mental health problems in Bangladeshi children: a preliminary study. Social Psychiatry and Psychiatric Epidemiology 36 (2), 94–99.

Nakamura, B.J., Ebesutani, C., Bernstein, A., Chorpita, B.F., 2009. A psychometric analysis of the Child Behavior Checklist DSM-oriented scales. Journal of Psychopathology and Behavioral Assessment 31, 178–189.

Palmieri, P.A., Smith, G.C., 2007. Examining the structural validity of the Strengths and Difficulties Questionnaire (SDQ) in a U.S. sample of custodial grandmothers. Psychological Assessment 19 (2), 189–198.

Pottick, K.J., Barber, C.C., Hansell, S., Coyne, L., 2001. Changing patterns of inpatient care for children and adolescents at the Menninger Clinic, 1988–1994. Journal of Counsseling and Clinical Psychology 69 (3), 573–577.

Reynolds, C.R., Livinstone, R.B., Wilson, V., 2008. Measurement and Assessment in Education, 2nd edition Pearson, New Jersey.

Rice, M.E., Harris, G.T., 2005. Comparing effect sizes in follow-up studies: ROC area, Cohen's *d*, and *r*. Law and Human Behavior 29 (5), 615–620.

Robins, L.N., Helzer, J.E., Croughan, J., Ratcliff, K.S., 1981. National Institute of Mental Health Diagnostic Interview Schedule: its history, characteristics and validity. Archives of General Psychiatry 38, 381–389.

Shaffer, D., Fisher, P., Lucas, C.P., Dulcan, M.K., Schwab-Stone, M.E., 2000. NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): description, differences from previous versions, and reliability of some common diagnoses. Journal of the American Academy of Child and Adolescent Psychiatry 39 (1), 28–38.

Sharp, C., Croudace, T.J., Goodyer, I.M., Amtmann, D., 2005. The Strengths and Difficulties Questionnaire: predicitive validity for help-seeking behaviour from parents over one year. Educational and Child Psychology 22 (3), 28–44.

Simpson, G.A., Bloom, B., Cohen, R.A., Blumberg, S., Bourdon, K.H., 2005. US children with emotional and behavioral difficulties: data from the 2001, 2002, and 2003. National Health Interview Surveys. Advance Data From Vital and Health Statistics, vol. 360, 1–16.

Stone, L.L., Otten, R., Engels, R.C.M.E., Verhulst, A.A., Janssens, J.M.A.M., 2010. Psychometric properties of the parent and teacher versions of the Strengths and Difficulties Questionnaire for 4- to 12-year olds: a review. Clinical Child and Family Psychology Review 13 (3), 254–274.

Swets, J.A., Pickett, R.M., 1982. Evaluation of Diagnostic Systems. Academic Press, Orlando.

Szatmari, P., Offord, D.R., Boyle, M.H., 1989. Ontario Child Health Study: prevalence of attention deficit disorder with hyperactivity. Journal of Child Psychology and Psychiatry 30 (2), 219–230.

Thurston, I.B., Phares, V., 2008. Mental health service utilization among African American and Caucasian mothers and fathers. Journal of Consulting and Clinical Psychology 76 (6), 1058–1067.

van Lang, N.D.J., Ferdinand, R.F., Oldehinkel, A.J., Ormel, J., Verhulst, F.C., 2005. Concurrent validity of the DSM-IV scales affective problems and anxiety problems of the Youth Self-Report. Behaviour Research and Therapy 43 (11), 1485–1494.

van Slyke, K.B., 2008. Assessing Childhood Difficulties: Comparing the SDQ and the BASC-2 (Dissertation Abstracts International). 68; p. 7289B (UMI No. AAT 3290039).

Verhulst, F.C., van der Ende, J., 1997. Factors associated with child mental health service use in the community. Journal of the American Academy of Child and Adolescent Psychiatry 36 (7), 901–909.

Vostanis, P., 2006. Strengths and Difficulties Questionnaire: research and clinical applications. Current Opinion in Psychiatry 19 (4), 367–372.

Vreugdenhil, C., van den Brink, W., Ferdinand, R., Wouters, L., Doreleijers, T., 2006. The ability of YSR scales to predict DSM/DISC-C psychiatric disorders among incarcerated male adolescents. European Child and Adolescent Psychiatry 15 (2), 88–96.

Zahner, G.E., Pawelkiewicz, W., DeFrancesco, J.J., Adnopoz, J., 1992. Children's mental health service needs and utilization patterns in an urban community: an epidemiological assessment. Journal of the American Academy of Child and Adolescent Psychiatry 31 (5), 951–960.