

# Psychometric Properties of the Parent and Teacher Versions of the Strengths and Difficulties Questionnaire for 4- to 12-Year-Olds: A Review

Lisanne L. Stone · Roy Otten ·  
Rutger C. M. E. Engels · Ad A. Vermulst ·  
Jan M. A. M. Janssens

Published online: 30 June 2010  
© The Author(s) 2010. This article is published with open access at Springerlink.com

**Abstract** Since its development, the Strengths and Difficulties Questionnaire (SDQ) has been widely used in both research and practice. The SDQ screens for positive and negative psychological attributes. This review aims to provide an overview of the psychometric properties of the SDQ for 4- to 12-year-olds. Results from 48 studies ( $N = 131,223$ ) on reliability and validity of the parent and teacher SDQ are summarized quantitatively and descriptively. Internal consistency, test–retest reliability, and inter-rater agreement are satisfactory for the parent and teacher versions. At subscale level, the reliability of the teacher version seemed stronger compared to that of the parent version. Concerning validity, 15 out of 18 studies confirmed the five-factor structure. Correlations with other measures of psychopathology as well as the screening ability of the SDQ are sufficient. This review shows that the psychometric properties of the SDQ are strong, particularly for the teacher version. For practice, this implies that the use of the SDQ as a screening instrument should be continued. Longitudinal research studies should investigate predictive validity. For both practice and research, we emphasize the use of a multi-informant approach.

**Keywords** Strengths and Difficulties Questionnaire · Reliability · Validity · Parent · Teacher

## Introduction

Many children's lives are troubled. Psychosocial childhood problems are common; research has shown that between 3 and 18% of all children suffer from some sort of psychopathology (Bourdon et al. 2005; Costello et al. 2003; Egger and Angold 2006; Ford et al. 2003; Meltzer et al. 2003; Zwirs et al. 2007). Behavioral disorders, such as oppositional defiant disorder (ODD), conduct disorder, and attention-deficit/hyperactivity disorder (ADHD), and emotional disorders, such as anxiety and depressive disorders are diagnosed most frequently in children (Canino et al. 2004; Egger and Angold 2006; Ford et al. 2003).

A substantial discrepancy has been found between the prevalence rates and the number of psychosocial problems being treated in childhood (see for a review Costello et al. 2005). One of the causes of this divergence may be the stigma (Corrigan 2004) associated with mental health care or limited access to care (Kataoka et al. 2002). Another explanation might be that psychosocial problems in the community are often not recognized or diagnosed (Costello et al. 2005). This is worrisome given the fact that problems in young children show relative stability over time (Caspi et al. 1996) and can potentially escalate or progress into psychiatric disorders. Thus, screening children at an early age for mental health problems and delivering early interventions, which might prevent these childhood problems from developing into more severe psychiatric disorders, is of great importance (Harrington et al. 1996). Though many instruments are available for screening children, The Child Behavior Check List (CBCL; Achenbach 1991) has long been viewed as the “gold standard” in assessing childhood problems. Recently, attention for early and quick detection of childhood psychopathology has increased. This has created room for other questionnaires

---

L. L. Stone (✉) · R. Otten · R. C. M. E. Engels ·  
A. A. Vermulst · J. M. A. M. Janssens  
Behavioural Science Institute, Radboud University Nijmegen,  
P.O. Box 9104, 6500 HE Nijmegen, The Netherlands  
e-mail: L.Stone@pwo.ru.nl

than the CBCL to be used as screening instruments. The launch of the Strengths and Difficulties Questionnaire (SDQ; Goodman 1997) has enabled researchers and clinicians to increase acceptability in respondents by offering a short and partly positively worded questionnaire (Goodman and Scott 1999). Whereas the CBCL is a very solid instrument in doing in-depth assessment, the SDQ may be more suitable for screening purposes. The SDQ is thus not a replacement of the CBCL by being the new gold standard but complements the field of childhood psychological assessment by adding a questionnaire which is shorter and quicker than the CBCL. The CBCL remains very useful though as an in-depth questionnaire. The SDQ has quickly become one of the most utilized *screening* instruments because it is able to measure both problem behavior and competencies at an early age. In the current study, we reviewed studies examining the psychometric properties of the parent and teacher versions of the SDQ.

The SDQ is a relatively short, user-friendly screening instrument of psychosocial problems for children, and worded more positively compared to other common questionnaires. Specifically, the SDQ has relatively few items (25 vs. 118) compared to the Child Behavior Check List (CBCL; Achenbach 1991). Another advantage of the SDQ is that it is free of charge and available online ([www.sdqinfo.com](http://www.sdqinfo.com)). The SDQ fits the current paradigm in the assessment of psychosocial problems, wherein the focus is expanded to include competencies or strengths in addition to assessing problems (Carr 2000; Rhee et al. 2001). The SDQ is based on the Rutter Questionnaires, which were developed in the 1960s (Rutter 1967). Goodman updated the items of the Rutter Questionnaires according to the current focus in child psychopathology, for example by adding items to concentration, peer relations, and social competence areas (Goodman 1994, 1997). The update is based on criteria from the Diagnostic and Statistical Manual of mental disorders, fourth edition (American Psychiatric Association 1994) and the International Classification of Diseases, tenth edition (World Health Organization 1992). Additionally, the instrument includes a prosocial scale, which was added to make the assessment more acceptable to respondents. Goodman (1994) devised items of the parent version of the prosocial scale, while the teacher version items were based on the Prosocial Behavior Questionnaire (PBQ; Weir and Duveen 1981). An impact supplement was added to the SDQ, enabling the informants to report on possible burden and distress (Goodman 1999).

The SDQ intends to measure both psychosocial problems and strengths (for example prosocial behavior) in children and youths aged 3–16 through a multi-informant approach. Parents and teachers can report difficulties and strengths among 3- to 16-year-olds, whereas youths aged 11–16 can report on their difficulties and strengths

themselves. The questionnaire consists of 25 items equally divided across five scales measuring emotional symptoms, conduct problems, hyperactivity-inattention, peer problems, and prosocial behavior. Except for the prosocial scale, the combined scale score reflects total difficulties, indicating the severity and the content of the psychosocial problems. The prosocial scale indicates the amount of prosocial characteristics a child shows (Goodman 1997).

The impact supplement comprises of eight questions. The first question asks whether the informant thinks the child has a problem, the remaining questions assess chronicity, distress, social impairment, and burden for others. From these questions, three dimensions can be inferred: perceived difficulties (is there a problem), impact score (distress and social incapacity on the child), and a burden rating (do symptoms impose a burden) (Goodman 1999).

As the SDQ is translated into over 60 languages, it has been widely used as a screening and research tool, a treatment-outcome measure, and a part of clinical assessment. In accordance with the increasing use of the SDQ, the body of research on the psychometric properties of the instrument is also growing substantially. Therefore, an overview of the results on psychometric properties, reliability, and validity would be very useful for researchers and practitioners.

The aim of this review is to review the psychometric properties of the parent and teacher versions of the SDQ for children aged 4–12 (primary school-aged children). Most research on the SDQ has focused on upper primary school-aged children and youngsters attending secondary school. Psychometric properties of the SDQ in these older children have been found sufficient in community (e.g., Koskelainen et al. 2001) and clinical samples (e.g., Becker et al. 2004), but research conducted on lower primary school-aged children shows mixed findings. Thus, it is important to review findings for primary school-aged children in order to draw conclusions about the suitability of the SDQ for younger children.

Having multiple informants reporting on the SDQ is valuable because psychosocial problems may be highly situational (Achenbach et al. 1987; Goodman et al. 2000c). Thus, the rater's perception of the situation may influence the ratings. Therefore, we have to investigate whether the psychometric properties of the SDQ in these informants differ and, based on the findings, examine possible implications for the use of the SDQ. Further, the utility of the SDQ is different in clinical versus community populations. In a clinical population, we assume the presence of psychosocial problems. Therefore, the SDQ should inform us about types of psychosocial problems, the duration, and perception of these problems. In a community population of children, we assume the presence of some but not all psychosocial problems; hence, the SDQ should be very

sensitive in detecting those children in the community who suffer from (developing) psychosocial problems. The aim of the SDQ is thus slightly different in clinical and community populations.

Specifically, we report results on internal consistency, test–retest reliability, and inter-rater agreement. As for validity, the results of construct, concurrent, capacity to discriminate, and predictive validity are reported.

## Methods

### Search Strategy and Selection for Identification of Studies

The electronic databases PsychINFO, PubMed, and ERIC were searched in March 2010 using the search terms “strengths and difficulties questionnaire,” “validity,” and “reliability.” Neither books nor unpublished articles were retrieved from the references.

Abstracts of selected studies were thoroughly read in order to determine whether they were potentially eligible for the inclusion in this review. Inclusion criteria were as follows:

- The target population had to be 4–12 years of age. The age was above the range in 27 out of 48 studies. Of those studies, 3.7% exceeded the age limit by 1 year, 7.4% by 2 years, 22.2% by 3 years, 25.9% by 4 years, 29.6% by 5 years, 7.4% by 6 years, and 3.7% by 7 years. Still, we included these studies in our review, as the results from younger children in those studies are important for our review. Whenever possible, only the results from primary school-aged children were extracted, and the results from secondary school-aged children were omitted.
- Studies had to assess the psychometric properties.
- Studies had to use the parent and/or teacher SDQ version but not self-report.
- Reports had to be available in English.

Eventually,  $k = 48$  studies were eligible for our review. All studies were published as articles in scientific journals. The publication dates of the 48 articles ranged from 1997 to March 2010. Methodological characteristics of each study are summarized in Table 1. The studies that were selected for this review are indicated with an asterisk in the reference list.

### Strategy for Analysis

The results of internal consistency (the extent to which items produce similar scores) (Cronbach 1951), test–retest reliability (the extent to which a questionnaire yields

similar results at different time points), and inter-rater agreement (the consensus between different raters) enabled us to report the outcomes systematically. In addition, a systematic comparison of the results of construct, concurrent, and capacity to discriminate was feasible. One of the most important assets of a questionnaire, the construct validity, here refers to the degree to which the SDQ is similar to other theoretical constructs of child psychopathology (Campbell and Fiske 1959). Concurrent validity is defined as the degree to which the SDQ scores relate to a theoretically similar construct, represented in a questionnaire. Capacity to discriminate refers to the ability of the SDQ to distinguish between groups that it should theoretically be able to distinguish between. Predictive validity is defined as the ability of the SDQ to predict scores on another criterion measure. As the method of examining predictive validity differs greatly with respect to research design, the results on predictive validity were not reviewed systematically but descriptively.

Reliability results were reported for each subscale as well as for the impact and total difficulties scales. Correlations were obtained and transformed first into Fisher’s  $Z$ -scores in order to enable the calculation of weighted correlations. The normally distributed Fisher’s  $Z$ -scores were weighted according to their sample size minus 3, and a weighted mean Fisher’s  $Z$ -score was computed by dividing the sum of the weighted Fisher’s  $Z$ -scores by the sum of their weights. The weighted mean  $Z$ -score was transformed back to a correlation coefficient  $r$  (Field 2001). Weighted mean correlations were reported separately by type of informant, parent, and teacher. Internal consistency values of  $\alpha = 0.70$  and below are generally considered low, values between  $\alpha = 0.70$  and  $\alpha = 0.80$  acceptable, and values of  $\alpha = 0.80$  and above good (Cohen 1977). Time intervals of test–retest reliability varied between 2 weeks and 6 months. Generally, test–retest correlations of  $r = 0.70$  and above are considered acceptable. Inter-rater agreement between parents and teachers was reported by subscale and total difficulties scale. No results on the impact scale were reported in the reviewed studies. As a rule of thumb, the meta-analytic mean of inter-rater agreement between parents and teachers ( $r = 0.27$ ) (Achenbach et al. 1987) is used as a benchmark of agreement or data quality (Goodman 2001). This meta-analytic mean was computed by extracting inter-rater agreement results from 41 studies on the CBCL. As the Achenbach et al., study is known as a landmark paper on inter-rater agreement, the use of 0.27 as a benchmark seems justified.

Item-level factor loadings were extracted from studies assessing construct validity. Factor loadings were not fully comparable due to the application of different extraction methods (like principal component analysis and principal axis factoring) and rotation methods (orthogonal or

**Table 1** Summary of studies included in the review

Nr Study	N	Age (years)	Informant	Source population	Country	General study aim	Psychometric properties assessed	Standard comparison diagnosis	Diagnostic domains assessed
1 Matsuiishi et al. (2008)	2,899	4–12	P	CO	Japan	Establish norms, factor structure	Internal consistency, construct validity	None	Total, prosocial
2 Perren et al. (2007)*	160	5–6	P, T	CO	Switzerland	Predictive value of prosocial behavior	Predictive validity	Berkeley Puppet Interview	Total, prosocial
3 Hill and Hughes (2007)	374	6	P, T	CL	United States (Texas)	Factor structure	Construct validity	Class play	Total, prosocial
4 Van Leeuwen et al. (2006)	3,169	4–8	P, T	CO	Belgium	Psychometric properties Dutch version	Internal consistency, inter-rater agreement, construct and concurrent validity	CBCL	Prosocial, hyperactivity, internalizing
5 Samad et al. (2005)	112	4–16	P	CL	Pakistan	Validity in Pakistan	Capacity to discriminate	ICD-10, Pediatric vs. Psychiatric	Total, prosocial
6 Bourdon et al. (2005)	9,878	4–17	P	CO	United States	Establish norms United States, evaluate SDQ	Internal consistency	Service use	Total, prosocial
7 Dickey and Blumberg (2004)	9,574	4–17	P	CO	United States	Establish factor structure in United States	Construct validity	None	Total, prosocial
8 Goodman et al. (2004)	1,028	5–17	P, T, S	CL	United Kingdom	Assess SDQ for improving detecting community psychiatric disorders	Capacity to discriminate	None	Total, prosocial
9 Becker et al. (2004)	543	5–17	P, T	CL	Germany	Examine reliability and validity of German SDQ	Internal consistency, construct, concurrent and capacity to discriminate	CBCL & clinical diagnoses (ICD-10)	Total, prosocial
10 Hawes and Dadds (2004)	1,359	4–9	P, T	CO	Australia	Psychometric properties Australian version	Internal consistency, concurrent validity	Diagnostic interviews	Total, prosocial
11 Malmberg et al. (2003)	493	5–15	P	CO, CL	Sweden	Validity of Swedish SDQ	Capacity to discriminate	Community versus Psychiatric	Total, prosocial
12 Glazebrook et al. (2002)	10,745	5–15	P	CO, CL	United Kingdom	Screen for utility of SDQ in pediatric clinics	Capacity to discriminate	Community versus Clinic	Total, prosocial
13 Mullick and Goodman (2001)	261	4–16	P, T, S	CO, CL	Bangladesh	Suitability of the SDQ for detecting psychiatric problems in Bangladesh	Capacity to discriminate	Community versus Clinic	Hyperkinesia, conduct disorder, emotional disorders or any psychiatric disorders
14 Goodman (2001)	10,438	5–15	P, T, S	CO	United Kingdom	Describe psychometric properties of SDQ	Internal consistency, inter-rater agreement, test retest reliability, construct and concurrent validity	DSM diagnoses	Total, prosocial
15 Klasen et al. (2000)	273	4–13	P	CO, CL	Germany	Compare SDQ with CBCL in Germany	Concurrent validity	CBCL and ICD diagnosis	Total, prosocial

Table 1 continued

Nr Study	N	Age (years)	Informant	Source population	Country	General study aim	Psychometric properties assessed	Standard comparison diagnosis	Diagnostic domains assessed
16 Goodman et al. (2000b)	7,984	5–15	P, T, S	CO	United Kingdom	SDQ as a means for improving detection of child psychiatric disorders in community	Capacity to discriminate	SDQ versus DAWBA	Total, prosocial
17 Goodman and Scott (1999)	132	4–7	P	CL	United Kingdom	Compare SDQ with CBCL in United Kingdom.	Concurrent and capacity to discriminate	CBCL	Total, prosocial
18 Goodman (1999)	699	5–15	P, T	CO, CL	United Kingdom	Validation study into extended SDQ version: impact scale	Capacity to discriminate	Community versus Clinic	Total, prosocial
19 Goodman (1997)	403	4–16	P, T	CO, CL	United Kingdom	Compare SDQ with Rutter scales	Concurrent and capacity to discriminate	Rutter questionnaires	Total, prosocial
20 Du et al. (2008)	1,965	3–17	P, T, S	CO, CL	China (Shanghai)	Assess norms, reliability, validity, factor structure	Internal consistency, inter-rater agreement, test retest reliability, construct, discriminant and capacity to discriminate	None	Total, prosocial
21 Van Roy et al. (2008)	32,914	10–19	P/proxy, S	CO	Norway	Construct validity of five-factor structure pre-, early, late adolescence	Construct validity	None	Total, prosocial
22 Parkes et al. (2008)	818	8–12	P	CL	Europe (8 regions)	Describe psychological symptoms in children with cerebral palsy & assess psychometric properties	Internal consistency	None	Total, prosocial
23 Zwirs et al. (2008)*	2,439	6–10	P, T	CO	The Netherlands (ethnically diverse)	Develop & validate screening instrument for externalizing disorders	Internal consistency, capacity to discriminate	None	Externalizing disorder
24 Mellor and Stokes (2007)	914	7–17	P, T, S	CO	Australia	Factor structure	Construct validity	None	Total, prosocial
25 Palmieri and Smith (2007)	733	4–17	P	CL	United States	Structural validity of P version for custodial grandmothers	Internal consistency, construct validity	None	Total, prosocial
26 Hysing et al. (2007)	1,040	7–9	P	CO	Norway	Evaluate sensitivity & specificity of SDQ in children with chronic illness	Concurrent validity	DAWBA	Total, prosocial
27 Becker et al. (2006)	1,459	6–18	P, T	CO	Europe (10 countries)	Examine SDQ P version in several European countries	Internal consistency, construct and concurrent validity	None	Total, prosocial

**Table 1** continued

Nr	Study	N	Age (years)	Informant	Source population	Country	General study aim	Psychometric properties assessed	Standard comparison diagnosis	Diagnostic domains assessed
28	Sharp et al. (2005)*	659	7–11	P, T	CO	United Kingdom	Establish predictive validity for P and T ratings	Predictive validity	Parental help-seeking behavior and worry	Total, prosocial
29	Kashala et al. (2005)	1,187	7–9	T	CO	Congo (Kinshasa)	Pilot SDQ, investigate mental health and association with school problems, etc.	Internal consistency, construct validity	None	Total, prosocial
30	Edmunds et al. (2005)	278	5	P	CO	United Kingdom (London)	Evaluate reliability & validity for SDQ in a health study among young children	Internal consistency, concurrent validity	Child Health Questionnaire	Total, prosocial
31	Mellor (2004)	917	7–17	P, T, S	CO	Australia	Reliability with younger respondents	Internal consistency, inter-rater agreement, test retest reliability	None	Total, prosocial
32	Widenfelt et al. (2003)	1,686	8–16	P, T, S	CO	The Netherlands	Psychometric properties of Dutch SDQ	Internal consistency, inter-rater agreement, concurrent validity	CBCL, YSR	Total, prosocial
33	Muris et al. (2003)	562	9–15	P, S	CO	The Netherlands	Examine psychometric properties	Internal consistency, test retest reliability, concurrent validity	CBCL, CDI, RCMAS, ADHDQ	Total, prosocial
34	Koskelainen et al. (2001)	735	7–16	P, T, S	CO	Finland	Evaluate psychometric properties	Internal consistency, inter-rater agreement, concurrent validity	CBCL, help-seeking questions	Total, prosocial
35	Goodman et al. (2007)	400	5–7	P	CO	United Kingdom	Check whether minor changes to questionnaire alter psychometrics	Capacity to discriminate	Rutter questionnaires	Total, prosocial
36	Smedje et al. (1999)	900	6–10	P	CO	Sweden	Validation study of Swedish SDQ	Internal consistency, construct validity	None	Total, prosocial
37	Kaptein et al. (2008)	967	6–12	P	CO, CL	The Netherlands	Assess differences in mental health of ID and non-ID children	Internal consistency, capacity to discriminate	None	Total, prosocial
38	Shojaei et al. (2008)	1,348	6–11	P	CO	France	Examine psychometric properties	Internal consistency	Socio-demographic data	Total, prosocial
39	Rothenberger et al. (2008)	2,406	7–16	P	CO	Germany	Examine psychometric properties of parent SDQ	Internal consistency, construct validity	None	Total, prosocial
40	Goodman et al. (2000b)	190	4–16	P, T, S	CL	United Kingdom and Bangladesh	Predict type of disorder from the SDQ	Concurrent validity	Clinical diagnosis	Total, prosocial

Table 1 continued

Nr Study	N	Age (years)	Informant	Source population	Country	General study aim	Psychometric properties assessed	Standard comparison diagnosis	Diagnostic domains assessed
41 Mathai et al. (2004)	130	4–14	P, T, S	CL	Australia	Examine agreement between SDQ and CAMHS	Concurrent and capacity to discriminate	Community child and adolescent mental health service (CAMHS)	Hyperactivity/inattention, conduct and emotional problems
42 Alyahri and Goodman (2006)	187	5–12	P, T	CO, CL	Yemen	Validate Arabic version of SDQ	Capacity to discriminate	DAWBA	Total, prosocial
43 Mathai et al. (2002)	130	4–14	P, T, S	CL	Australia	Examine usefulness of the SDQ as a screening measure	Inter-rater agreement and concurrent validity	CAMHS	Total, prosocial
44 Sanne et al. (2009)	8,999	7–9	P, T	CO	Norway	Test five-factor model	Internal consistency, construct validity	None	Total, prosocial
45 Lai et al. (2009)	3,722	6–12	P, T	CO, CL	China	Validate Chinese version of the SDQ	Internal consistency, test–retest reliability, capacity to discriminate	None	Total, prosocial
46 Janssens and Deboutte (2009)	292	3–18	P, caregiver, S	CL	Belgium	Usefulness of SDQ for screening children entering welfare	Internal consistency, concurrent validity	CBCL	Total, prosocial
47 Syed et al. (2009)	556	5–11	P	CO	Pakistan	Compare SDQ with CBCL	Concurrent validity	CBCL	Total, prosocial
48 Vogels et al. (2009)	2,066	7–12	P, S	CO	The Netherlands	Compare three questionnaires	Capacity to discriminate	CBCL	Total, prosocial

Studies indicated with an asterisk employed a longitudinal design

P parent, T teacher, S self-report, CO community, CL clinical

oblique) in studies using exploratory factor analysis. The estimation methods were different (maximum likelihood or weighted least squares) in studies using confirmative factor analysis. To gain insight into the quality of the measurement model of the SDQ, loadings were categorized into low (<0.40), medium ( $\geq 0.40$ – $\leq 0.70$ ), or high (>0.70). Also, weighted mean factor loadings were calculated on item level.

Concurrent validity was reported mainly as the correlation of SDQ measures with measures of psychopathology like the CBCL or other measures of psychopathology. In the reviewed studies that examined capacity to discriminate, receiver operating characteristic (ROC) analyses were conducted to distinguish between high- and low-risk samples, generating the area under curve (AUC). An AUC with a value of 1 shows perfect capacity to discriminate and a value of 0.5 the absence of capacity to discriminate. Sensitivity (i.e., the proportion of children who are correctly identified by the SDQ as having psychosocial problems) and specificity (i.e., the proportion of children who are correctly identified by the SDQ as not having psychosocial problems) results were extracted and summarized. Again, the results were weighted according to their sample size.

Due to unique research designs in some studies, not all results could be captured in tables. Results from these studies are reported descriptively, as are the results on predictive validity.

## Results

### Internal Consistency

Weighted mean and the range of unweighted internal consistency reliability estimates by type of informant are presented in Table 2, as extracted from 26 studies. Prosocial behavior, emotional symptoms, conduct problems, and peer problems showed internal consistencies below 0.70 for parents. Teacher ratings showed higher internal consistencies with only peer problems having a value below 0.70.

### Test–Retest Reliability

Weighted correlations and the range of unweighted correlations from six studies are presented in Table 3. At the subscale level as well as for the impact scale, parent ratings tended to be less reliable over time compared to teacher ratings.<sup>1</sup>

<sup>1</sup> All reported results from this section forward are significant at the  $p < 0.05$  level.

**Table 2** Weighted mean internal consistency results on the SDQ specified by informant

	Informant			
	Parent	Range	Teacher	Range
Prosocial behavior	0.67	0.54–0.84	0.82	0.79–0.86
Hyperactivity/inattention	0.76	0.58–0.85	0.83	0.66–0.89
Emotional symptoms	0.66	0.60–0.76	0.73	0.63–0.80
Conduct problems	0.58	0.46–0.76	0.70	0.63–0.84
Peer problems	0.53	0.30–0.76	0.63	0.35–0.77
Total difficulties	0.80	0.53–0.84	0.82	0.62–0.85
Impact scores	0.81	0.69–0.87	0.85	–
<i>N</i>	53,691	–	21,866	–

*Note:* Results on internal consistency retrieved from the following studies: Becker et al. (2004, 2006), Bourdon et al. (2005), Du et al. (2008), Edmunds et al. (2005), Goodman (2001), Hill and Hughes (2007), Hawes and Dadds (2004), Janssens and Deboutte (2009), Kaptein et al. (2008), Kashala et al. (2005), Koskelainen et al. (2001), Lai et al. (2009), Malmberg et al. (2003), Matsuishi et al. (2008), Muris et al. (2003), Parkes et al. (2008), Perren et al. (2007), Rothenberger et al. (2008), Sanne et al. (2009), Shojaei et al. (2008), Smedje et al. (1999), Van Leeuwen et al. (2006), Van Roy et al. (2008), Vogels et al. (2009), and Widenfelt et al. (2003).  $k = 26$

**Table 3** Weighted mean test–retest correlations on the SDQ specified by informant

	Informant			
	Parent	Range	Teacher	Range
Prosocial behavior	0.65	0.43–0.78	0.79	0.50–0.84
Hyperactivity/inattention	0.71	0.48–0.85	0.85	0.64–0.89
Emotional symptoms	0.66	0.47–0.82	0.72	0.40–0.80
Conduct problems	0.66	0.52–0.89	0.77	0.58–0.86
Peer problems	0.66	0.61–0.91	0.77	0.58–0.82
Total difficulties	0.76	0.72–0.86	0.84	0.55–0.90
Impact scores	0.57	–	0.68	–
<i>N</i>	2,852	–	1,693	–

*Note:* Results on test–retest reliability retrieved from the following studies: Du et al. (2008), Goodman (1999, 2001), Lai et al. (2009), Mellor (2004), and Muris et al. (2003).  $k = 6$

### Inter-Rater Agreement

The results of parent and teacher inter-rater agreement correlations from eight studies by weighted mean correlations and by the range of unweighted correlations are presented in Table 4. The weighted mean correlations varied between 0.26 and 0.47. All subscales, except the prosocial scale, had a higher mean than the meta-analytic mean of 0.27.



**Table 4** Weighted parent and teacher inter-rater agreement correlations on the SDQ

	Weighted total	Range
Prosocial behavior	0.26	0.22–0.30
Hyperactivity/inattention	0.47	0.44–0.61
Emotional symptoms	0.28	0.23–0.41
Conduct problems	0.34	0.27–0.65
Peer problems	0.35	0.27–0.59
Total difficulties	0.44	0.37–0.62
Impact scores	–	–
<i>N</i>	14,811	

*Note:* Results on inter-rater agreement retrieved from the following studies: Du et al. (2008), Goodman (1997, 2001), Koskelainen et al. (2001), Mathai et al. (2002), Mellor (2004), Van Leeuwen et al. (2006), and Widenfelt et al. (2003).  $k = 8$

### Construct Validity

A review of the results of the five-factor structure for children aged 4–12 is presented in Table 5. In the parent version, the number of factor loadings was summed across 13 studies. Of these 13 studies, six studies examined also the teacher version. It should be noted that Smedje et al. (1999) and Hawes and Dadds (2004) split their sample into boys and girls, each study generating two sets of factor loadings. Sanne et al. (2009) applied both EFA and CFA, which also generated two sets of factor loadings. Therefore, factor loadings for the parent version summed to 16.

For parent and teacher versions, most items showed satisfactory factor loadings  $>0.40$ – $\leq 0.70$ . For the parent version, highest loadings were found on the hyperactivity-inattention subscale and lowest on the conduct problems subscale. For teachers, highest loadings were found on the prosocial subscale and lowest on the peer problems scale. However, in 11 out of 14 studies, the results of these factor analyses were obtained by conducting exploratory factor analysis (EFA).

Eight studies applied confirmatory factor analysis; however, only four are presented in Table 5 (Palmieri and Smith 2007; Van Leeuwen et al. 2006; Van Roy et al. 2008; Sanne et al. 2009) because four out of the total of eight studies did not report factor loadings (Becker et al. 2004; Dickey and Blumberg 2004; Hill and Hughes 2007; Mellor and Stokes 2007).

These eight studies are discussed below. Dickey and Blumberg (2004) found support for a three-factor structure of prosocial, externalizing, and internalizing problems. Van Leeuwen et al. (2006) examined a five-factor model and a three-factor model in two samples. Support was found for the five-factor model for the parent and teacher versions. The three-factor model for the parent and teacher versions revealed a worse model fit. The findings of Becker et al.

(2004), Van Roy et al. (2008), and Sanne et al. (2009) provided support for the five-factor model for both the parent and teacher versions, but this factor structure was not found by Mellor and Stokes (2007) and was only marginally adequate in Hill and Hughes' (2007) study. Palmieri and Smith (2007) confirmed the five-factor structure for custodial grandparents.

### Concurrent Validity

Regarding results of concurrent validity, weighted SDQ-CBCL correlations and the range of unweighted correlations are presented in Table 6. The presented correlations do not include all CBCL subscales. In the majority of the reviewed studies, SDQ problem scales correlated with the CBCL subscales that covered similar concepts in general, that is, externalizing, attention problems, internalizing, and social problems. Weighted correlations of 0.76 for both parent (range of unweighted  $r = 0.70$ – $0.87$ ) and teacher ratings (range of unweighted  $r = 0.68$ – $0.87$ ) were found between the SDQ total difficulties and CBCL total scales. At the subscale level, conduct problems, externalizing and hyperactivity, and attention problems correlated sufficiently, while emotional symptoms, internalizing and peer problems, and social problems showed correlations below 0.70. The SDQ impact scale and CBCL total scale correlated below 0.70.

### *SDQ Correlations with Measures of General Psychopathology*

The SDQ has correlated with other measures of general psychopathology. High correlations have been found between SDQ total difficulties and Rutter total deviance scales for parent ( $r = 0.88$ ) and teacher ( $r = 0.92$ ) ratings (Goodman 1997). Another study replicated the correlation between SDQ total difficulties and Rutter total deviance scales for parent ratings ( $r = 0.76$ ) (Goodman et al. 2007). Somewhat lower correlations were found between the parent-rated SDQ and the Chinese version of the parent-rated Conner's Parent Symptom Questionnaire (PSQ; Du et al. 1995), SDQ total difficulties and PSQ total score had  $r = 0.63$ , conduct problems and conduct problems had  $r = 0.53$ , hyperactivity-inattention and impulsivity-hyperactivity had  $r = 0.56$ , hyperactivity-inattention and hyperactivity index score had  $r = 0.61$ , and hyperactivity-inattention and learning problems had  $r = 0.58$  (Du et al. 2008). The Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA; Gowers et al. 1999), a clinician-based mental health assessment tool, has been correlated with the SDQ total difficulties, resulting in moderate correlations for parent  $r = 0.38$  and teacher  $r = 0.46$  ratings. At the subscale level, correlations

**Table 5** Frequencies of factor loadings on item level of the SDQ specified by informant

	Frequencies							
	Parent <i>N</i> = 43,274 (13 studies)				Teacher <i>N</i> = 19,105 (6 studies)			
	<.40	≥.40–<.70	≥.70	<i>M</i>	<.40	≥.40–<.70	≥.70	<i>M</i>
<b>Prosocial behavior</b>								
1 Considerate	0	12	4	0.65	0	5	2	0.70
4 Shares	1	15	0	0.56	0	2	5	0.71
9 Caring	0	9	7	0.68	1	0	6	0.80
17 Kind to kids	1	14	1	0.60	0	2	5	0.74
20 Helps out	1	9	4	0.63	0	1	6	0.76
<b>Hyperactivity/inattention</b>								
2 Restless	1	8	7	0.63	0	1	5	0.80
10 Fidgety	0	8	8	0.60	0	2	4	0.81
15 Distractible	1	4	11	0.74	0	2	5	0.77
21 Reflective*	1	14	1	0.57	0	6	1	0.56
25 Persistent*	1	6	9	0.70	0	2	5	0.73
<b>Emotional symptoms</b>								
3 Somatic complaints	4	10	1	0.47	1	6	0	0.48
8 Worries	0	7	9	0.70	0	3	4	0.73
13 Unhappy	0	14	2	0.63	0	6	1	0.65
16 Clingy	0	14	2	0.65	0	4	3	0.73
24 Fears	0	11	5	0.66	0	3	4	0.80
<b>Conduct problems</b>								
5 Tempers	3	12	1	0.52	0	5	1	0.67
7 Obedient*	3	11	1	0.46	2	0	2	0.43
12 Fights	1	11	4	0.61	0	3	4	0.72
18 Lies, cheats	3	10	2	0.62	1	4	2	0.63
22 Steals	2	12	1	0.56	0	6	0	0.59
<b>Peer problems</b>								
6 Solitary	3	12	1	0.61	2	2	2	0.67
11 Good friend*	5	9	1	0.52	1	3	2	0.64
14 Popular*	0	11	3	0.67	1	2	2	0.58
9 Picked on, bullied	1	12	2	0.58	1	4	1	0.51
23 Best with adults	2	9	3	0.63	2	2	2	0.68

*Note:* Results on construct validity retrieved from the following studies: Becker et al. (2006), Dickey and Blumberg (2004), Du et al. (2008), Goodman (2001), Hawes and Dadds (2004), Hill and Hughes (2007), Matsushita et al. (2008), Muris et al. (2003), Kashala et al. (2005), Palmieri and Smith (2007), Sanne et al. (2009), Smedje et al. (1999), Van Leeuwen et al. (2006), and Van Roy et al. (2008). *k* = 14. Items indicated with an asterisk are reversed items

between HoNOSCA and the hyperactivity-inattention scales of  $r = 0.33$  for parent and  $r = 0.41$  for teacher ratings have been reported (Mathai et al. 2002).

*SDQ Correlations with Measures of Specific Psychopathology*

The parent-rated SDQ correlated with the clinician-rated ADHD-RS-IV (DuPaul et al. 1998) in that total difficulties, and total score had  $r = 0.50$ . At the subscale level, hyperactivity-inattention and hyperactivity-impulsivity had  $r = 0.54$ . The SDQ prosocial scale correlated with the

parent-rated Child Health and Illness Profile-Child Edition (CHIP-CE; Riley et al. 2004) on the subscales of resilience  $r = 0.41$  and risk avoidance  $r = 0.40$  (Becker et al. 2006). The SDQ also correlated with the parent-rated ADHDQ-P (Scholte and Van der Ploeg 1998) on total difficulties with total score  $r = 0.67$ , hyperactivity-inattention with total score  $r = 0.73$ , and at the subscale level on hyperactivity-inattention with attention-deficit  $r = 0.65$ , and with hyperactivity  $r = 0.72$ . Correlations have been found between the parent-rated SDQ and the parent-rated Child Depression Inventory (CDI-P; Kovacs 1981), in that total difficulties and total score had  $r = 0.73$  and emotional

**Table 6** Concurrent validity: weighted SDQ-CBCL correlations specified by informant

	Informant			
	Parent	Range	Teacher	Range
Conduct problems/ externalizing	0.71	0.60–0.84	0.79	0.74–0.86
Hyperactivity/attention problems	0.69	0.64–0.78	0.77	0.76–0.80
Emotional symptoms/ internalizing	0.64	0.44–0.77	0.58	0.40–0.80
Peer problems/social problems	0.52	0.41–0.75	0.57	0.48–0.71
Total/total	0.76	0.70–0.87	0.76	0.68–0.87
Impact/total	0.46	0.44–.051	0.53	–
N	4,590	–	784	–

Note: Results on concurrent validity are retrieved from the following studies: Becker et al. (2004), Janssens and Deboutte (2009), Klasen et al. (2000), Koskelainen et al. (2001), Goodman and Scott (1999), Muris et al. (2003), Syed et al. (2009), Van Leeuwen et al. (2006), and Widenfelt et al. (2003).  $k = 9$

symptoms and total score had  $r = 0.67$ . The parent-rated SDQ correlated with the parent-rated Revised Children's Manifest Anxiety Scale (RCMAS-P; Reynolds and Richmond 1978), in that difficulties and total anxiety score had  $r = 0.72$ , and emotional symptoms and total anxiety score had  $r = 0.73$  (Muris et al. 2003).

#### Associations of the SDQ with the DAWBA, DMS-IV Diagnoses, and Risk Factors in Community Samples

An SDQ algorithm was developed in order to predict whether any psychiatric disorder is “unlikely,” “possible,” or “probable” (Goodman et al. 2000b). With this algorithm, children with a psychiatric diagnosis, as identified by the Development and Well-Being Assessment (DAWBA; Goodman et al. 2000b), were correctly classified as probably having a disorder in 77.3% of the cases. Using the SDQ algorithm, out of the children who were identified as having hyperactivity or conduct-oppositional or emotional disorder diagnosis according to DAWBA, 91% were rated as probable for a hyperactivity disorder, 60% were rated as probable for a conduct-oppositional disorder, and 44% were rated as probable for an emotional disorder (Hysing et al. 2007).

The SDQ algorithm was used in a study to generate diagnoses from SDQ scores. These diagnoses were compared with diagnoses given by independent clinicians or clinical teams based on *DSM-IV* (1994) criteria. Agreement (expressed in the rank-order correlation tau) between SDQ generated and clinical team diagnoses was found for hyperactivity ( $\tau = 0.44$ ), and conduct ( $\tau = 0.56$ ) and emotional ( $\tau = 0.39$ ) disorders. Reasonable correlations

were found between SDQ generated and independent clinician diagnoses for hyperactivity ( $\tau = 0.43$ ), and conduct ( $\tau = 0.30$ ) and emotional ( $\tau = 0.26$ ) disorders (Mathai et al. 2004).

Prevalence of *DSM-IV* (1994) diagnoses of high- (extreme 10% of sample) versus low-risk (90% of sample) groups based on parent- and teacher-rated SDQ scores differed. SDQ scores were compared with clinical diagnoses, which were assigned based on the DAWBA. Differences in prevalence between high- and low-risk groups showed that all (sub)scales were associated with *DSM-IV* diagnoses. The odds ratio (OR) for having a psychiatric disorder in the high-risk group was 15.7 for parent- and 15.2 for teacher-rated SDQs, across the total difficulties scale and the subscales (Goodman 2001).

A similar study assessed children with the Diagnostic Interview Schedule for Children, Adolescents, and Parents (DISCAP; Holland and Dadds 1995) and subsequently assigned *DSM-IV* diagnoses. Significant differences were found between high- and low-risk groups on each SDQ subscale and the total difficulties scale, indicating that higher scores are associated with a greater probability of being assigned a *DSM-IV* diagnosis. The odds ratio for having a psychiatric disorder in the high-risk group was 11.7 based on total difficulties and 14.9 based on the impact scale. In addition, severity of psychosocial problems was rated by clinicians and correlated with parent-rated SDQ scores for total difficulties ( $r = 0.47$ ) and the impact scale ( $r = 0.57$ ) (Hawes and Dadds 2004).

Risk factors such as having contact with a mental health professional or general practitioner (GP), attending special education, or having a desire of using these type of services but not being able to afford them have been shown to be associated with high parent-rated SDQ scores. Learning disability, ADHD, declining health, and demographic variables, such as living below the poverty line, living in single-parent, or reconstituted families, were significantly associated with high parent rated SDQ scores (Bourdon et al. 2005). For 26 children, parent-rated SDQ total difficulties were associated with (consideration of) service use (OR = 8.7) (Koskelainen et al. 2001). Parent-rated SDQ total difficulties ( $r = 0.16$ ), emotional symptoms ( $r = 0.15$ ), and peer problems ( $r = 0.15$ ) were associated with additional service use in 68 children receiving care in a welfare institution. Further, the need for additional help was predicted by the impact score of parents (OR = 1.37) and caregivers (OR = 1.50) but not by their total difficulties scores (OR = 1.07, OR = 1.03) (Janssens and Deboutte 2009).

#### Capacity to Discriminate

In Table 7, weighted AUC values are presented by informant. The combined AUC represents a weighted average

**Table 7** Weighted area under curves (by SE) on the SDQ specified by informant

	Informant			
	Parent	Range	Teacher	Range
Prosocial behavior	0.71	0.39–0.82	0.65	0.64–0.67
Hyperactivity/inattention	0.90	0.76–0.97	0.95	0.90–0.95
Emotional symptoms	0.79	0.69–0.85	0.84	0.65–0.88
Conduct problems	0.92	0.68–0.97	0.86	0.82–0.87
Peer problems	0.71	0.49–0.78	0.57	0.45–0.69
Total difficulties	0.87	0.64–0.91	0.83	0.65–0.91
Impact scores	0.86	0.83–0.87	0.88	0.85–0.89

*Note:* Results on capacity to discriminate are retrieved from the following studies: Alyahri and Goodman (2006), Du et al. (2008), Goodman (1997), Goodman and Scott (1999), Klasen et al. (2000), Malmberg et al. (2003), Mullick and Goodman (2001), and Samad et al. (2005).  $k = 8$

of the AUC in each study. The AUCs were weighted by their standard error. For the subscales, prosocial behavior, and peer problems, the AUC values were just above 0.5, indicating that, for teacher ratings, the ability of these subscales to distinguish between children with diagnoses, and those without, is just above chance level. For the remaining scales, AUC values are satisfactory.

Two studies could not be incorporated in Table 7 because standard errors or upper bounds were not given. Becker et al. (2004) report AUCs for the total difficulties (0.77, 0.75), emotional symptoms (0.69, 0.65), conduct problems (0.81, 0.82), and hyperactivity-inattention (0.77, 0.80) scales for the parent and teacher versions, respectively. So, except for the emotional symptoms scale, the SDQ is adequately able to differentiate between children with and without clinical diagnoses. In a study by Lai et al. (2009), AUC values were reported for emotional symptoms (0.79, 0.70), conduct problems (0.89, 0.86), hyperactivity-inattention (0.86, 0.85), peer problems (0.71, 0.69), prosocial behavior (0.60, 0.69), and total difficulties (0.84, 0.78), for the parent and teacher versions.

Samad et al. (2005) and Malmberg et al. (2003) assessed sensitivity and specificity of the parent-rated total difficulties and impact scales. The percentages of children identified by the SDQ as having a psychiatric disorder and who did have a disorder (true positives) were 69 and 82.4% for total difficulties, and respectively 66 and 82.7% for the impact scale (true positives). Children who did not have a psychiatric disorder were correctly identified as such (true negatives) 71 and 85.4% of the time by total difficulties, and 86 and 87.8% of the time by the impact scale. At the subscale level, sensitivity ranged from 56.6 to 75% and specificity from 66 to 88.1%. Two other studies assessed

sensitivity and specificity by combining parent and teacher reports only for the hyperactivity-inattention and emotional and conduct problems subscales. Goodman et al. (2000b) found sensitivity to be 89, 81, and 90%, respectively, on the aforementioned subscales in a London sample and 89, 86, and 86% in a Dhaka sample. Reported specificity values were 78, 80, and 47% in the London sample and 81, 84, and 82% in the Dhaka sample. Mathai et al. (2004) reported sensitivity of 44% for the hyperactivity-inattention scale, indicating that 44% of children with ADHD symptoms were correctly identified by the scale as such. Children presenting with emotional symptoms were correctly identified as having emotional symptoms in 36% of the cases. The scale conduct problems identified 93% of the children showing conduct problems correctly. So, the proportion of true positives that are correctly identified by the SDQ was higher for the conduct problems scale, than it was for the hyperactivity-inattention and emotional symptoms scale.

Goodman et al. (2000a) and Goodman et al. (2004) tested sensitivity in a community and clinical samples. Combined parent (or caregiver) and teacher reports yielded sensitivity of 62.1 and 82.2% in detecting any psychiatric disorder, respectively, in the community and clinical samples. When only parent report was used, sensitivity dropped to 29.8% in the community sample and to 51.4% in the clinical sample. For teacher reports only, sensitivity dropped to 34.5 and 59.8% in the community and clinical samples, respectively. Sensitivity for detecting conduct-oppositional, hyperkinetic, ADHD, anxiety, depressive, as well as less common disorders was also assessed. Results were comparable to sensitivity found in detecting any other psychiatric disorder, except for detecting anxiety disorder in the community. Sensitivity was only 45.5% for parent and teacher reports combined and even lower for teacher report only, with a detection rate of 15.9%. Parent report correctly identified anxiety disorders 33.8% of the time, a significant difference to teacher report.

When comparing children with and without intellectual disability (ID), 60.9% with ID were found to have an elevated SDQ score compared to 9.8% of children without ID (Kaptein et al. 2008). A somewhat similar result was obtained for children with chronic illness (CI); 20% of them scored high based on parent-rated SDQ total difficulties, while 11% of children who did not have CI scored high (Hysing et al. 2007). Children attending pediatric outpatient clinics were more than twice as likely to score in the abnormal SDQ range compared to children from the community (OR = 2.33). The chance of scoring in the abnormal range was even greater for children attending a pediatric clinic for brain disorder (OR = 5.8) compared to community children (Glazebrook et al. 2002).

Goodman (1999) directed special attention to the impact scale of the SDQ. The three concepts of the impact scale, perceived difficulties, impact score, and burden rating, showed a different distribution in community and clinical samples ( $\chi^2 = 67.8$ ), confirming the idea that problems of children in the community sample are not perceived as severe as problems of children in the clinical sample. Lastly, SDQ scores differed according to treatment status. Children currently receiving treatment for psychosocial problems had higher SDQ scores ( $M = 15.0$ ) compared to children not receiving treatment ( $M = 8.0$ ) (Hawes and Dadds 2004).

### Predictive Validity

Evidence for the predictive validity of the SDQ has been found in three studies. The first focused on the stability of parent ratings, the second on help-seeking behaviors, and the third on prosocial behavior. Hawes and Dadds (2004) found that SDQ scores remained relatively stable over a 12-month period for the total difficulties  $r = 0.77$  and impact  $r = 0.63$  scales. For the subscales, comparable correlations were found for hyperactivity-inattention,  $r = 0.77$ , prosocial,  $r = 0.64$ , conduct,  $r = 0.65$ , emotional,  $r = 0.71$ , and peer problems  $r = 0.61$ .

Sharp et al. (2005) found that, over 1 year, parent- and teacher-rated SDQ scores predicted parental help-seeking behaviors and worry about the child. Over three time points (6-month intervals), parent-rated emotional problems were associated with seeking help from family (OR = 1.09). Parent-rated total difficulties at 12 months were associated with worries (OR = 1.06). Emotional problems rated by parents at baseline and 6 months, predicted worries (OR = 0.85; OR = 1.33). Teacher-rated baseline total difficulties scores were associated with seeking help from a GP (OR = 0.17) and from a friend (OR = 14.88). The rate of change in total difficulties rated by teachers was associated with seeking help from school (OR = 1.13) and GP (OR = 1.25). Teacher-rated total difficulties at 6 months were associated with parental worry (OR = 1.12). Peer problems rated by teachers were associated with parental worry 6 months later (OR = 1.57).

Perren et al. (2007) examined the role of prosocial behavior in kindergarten longitudinally. In addition to parent and teacher SDQ ratings, children were able to perform as informants regarding their problems by using the Berkeley Puppet Interview (BPI; Measelle et al. 1998). Emotional symptoms, conduct problems, and hyperactivity-inattention at age five predicted subsequent emotional symptoms, conduct problems, and hyperactivity-inattention, as rated by multiple informants (i.e., parents, teachers, and children) at age six ( $\beta = 0.530$ ;  $\beta = 0.500$ ;  $\beta = 0.667$ , respectively). The level of prosocial behavior, in combination with the level of emotional

symptoms at age five, predicted emotional symptoms at age six. Children showing high levels of prosocial behavior and high levels of emotional problems at age five showed the highest level of emotional symptoms at age six, but children exhibiting high levels of prosocial behavior and low levels of emotional symptoms at age five showed the lowest levels of emotional symptoms at age six.

### Discussion

The aim of this review was to contribute to a better understanding of the psychometric properties of the SDQ. A total of 48 studies were reviewed. Several indications for research and practice regarding reliability and validity of the SDQ follow from this review.

#### Internal Consistency

Results from an impressive number of studies show acceptable internal consistency for the total difficulties and impact scale for both parent and teacher ratings. At the subscale level, we found differences between parent and teacher ratings. Except for hyperactivity-inattention scale, which had an adequate internal consistency, the prosocial scale, emotional, conduct, and peer problems scales showed only moderate internal consistencies for parent ratings. For teacher ratings, the peer problems scale showed a moderate alpha, while the remaining scales showed adequate internal consistency. The items of the peer problems scale may not reflect the same construct, as alphas for this scale are lowest for parent and teacher versions. The only item measuring problem behavior is, in our opinion, “picked on or bullied by other children”. Remaining items seem to reflect loneliness on the one hand (rather solitary, tends to play alone; has at least one good friend) and sociability on the other (generally liked by other children; gets on better with adults than with other children).

An explanation for the difference in internal consistency between parents and teachers is that for parents, the items from the subscales may be less one-dimensional than for teachers, which may refer to a halo effect for teachers (Abikoff et al. 1993; Nisbett and Wilson 1977). Halo effects occur when one class of behavior influences the perception, and thus the rating, of other behaviors. Specifically, halo effects have been found to influence ratings of ADHD and ODD (Abikoff et al. 1993; Jackson and King 2004).

#### Test–Retest Reliability

The parent version of the SDQ had lower reliability over time compared to the teacher version, specifically at the

subscale level. All parent-rated subscales, except the hyperactivity-inattention subscale, showed correlations below  $r = 0.70$ , whereas teacher subscales were all above  $r = 0.70$ . The total difficulties scales for parent and teacher ratings showed good test–retest reliability. Only the impact scale showed to be less reliable over time. The moderate over-time correlation for the impact scale may be due to the time interval of 4–6 months that was used in the study assessing the impact scale (Goodman 2001), in contrast to the time interval of 2 weeks to 6 months used in studies assessing the total difficulties scale (Du et al. 2008; Goodman 1999; Goodman 2001; Lai et al. 2009; Mellor 2004; Muris et al. 2003). The difference in parent versus teacher ratings at the subscale level may be explained in that parents are more prone to detect changes in their child's mood, as they usually spend more time with their child than their teacher does. This may have caused the correlation to be lower for parent than for teacher ratings.

#### Inter-Rater Agreement

Compared to the average inter-rater agreement reported for other measures of child psychopathology, the inter-rater agreement between parent and teacher ratings for total scales and subscales was predominantly better (Achenbach et al. 1987). However, reliability remains modest, which is a well-known phenomenon in psychological assessment. Although inter-rater agreement is valuable to test whether children behave similarly across situations, its use may be less valuable as a psychometric property.

#### Construct Validity

In five studies, the proposed five-factor structure was supported for both parent and teacher versions using confirmatory factor analysis. Recently, support was found for the five-factor model for the parent and teacher versions in a very large sample (Sanne et al. 2009). Only one study (Dickey and Blumberg 2004) found more support for a three-factor structure (internalizing, externalizing, and prosocial behavior) for the parent version. An explanation for the difference in factor structure between the studies of Dickey and Blumberg and Becker et al. (2004), which tested only the parent version using CFA, might be cross-cultural inequivalence (Berry et al. 2002). Parents from the United States may perceive problems differently than German parents do, which could lead to inconsistencies in factor structures.

In this review, most evidence was thus found for the original five-factor structure of prosocial behavior, hyperactivity/inattention, conduct, emotional, and peer problems. An important methodological aspect of construct validity needs to be highlighted. Despite the theoretical

foundation for a five-factor structure, non-normal distribution of scores, and a three-item response category, most studies reported results of exploratory factor analysis and principal component analysis. Both techniques are not suited to test the underlying structure of the SDQ. As the SDQ is based on theoretical constructs concerning child psychopathology (Goodman 1997), scores are non-normally distributed, and the response category is limited; therefore, confirmatory factor analysis (CFA) should be the first method of choice when investigating factor structure (Sanne et al. 2009).

#### Concurrent Validity

Many studies, comparable in some but not all cases, have validated the SDQ. Summarizing and interpreting the results from these studies is therefore complex. Correlations between SDQ and CBCL scales showed to be high for both parent and teacher ratings at the total scales level. The SDQ is thought to measure the same constructs as the CBCL, and these high correlations support that notion. However, at the subscale level, evidence for concurrent validity is less clear. The SDQ emotional and peer problems scales correlated moderately with the CBCL internalizing and social problems scales for both parent and teacher ratings. Further inspection of the CBCL internalizing subscales showed that the CBCL Anxious/Depressed subscale is very well represented by providing three out of five items which are very comparable with the items from the SDQ Emotional Symptoms subscale. However, no items from the CBCL Withdrawn subscale and only one from the Somatic Complaints and Emotionally Reactive subscales are represented in the SDQ Emotional Symptoms subscale. The Withdrawn subscale consists of items that reflect the autism spectrum disorders (ASD), which are not included in the SDQ. The overlap between the CBCL internalizing subscales and the SDQ Emotional Symptoms scale is thus quite small, which may explain the moderate correlation found in our review.

The SDQ impact correlated moderately with the CBCL total problems scale for both parent and teacher ratings. Experience of social impairment and substantial distress caused by psychiatric symptoms is nowadays a part of the diagnostic criteria for a psychiatric disorder (American Psychiatric Association 1994; World Health Organization 1992). The CBCL does not contain social impairment and distress items that would be similar to the SDQ impact supplement. Hence, the moderate correlation between the SDQ impact and CBCL total problems scales may indicate that these scales are conceptually different. The impact scale also correlated with a parental burden scale resulting in  $r = 0.74$  (Goodman 1999). This parental burden scale is thought to be more comparable to the impact scale than is

the CBCL total scale. The CBCL total scale focuses on symptoms of psychosocial problems, whereas the impact and parental burden scale focuses on the perception of the consequences of psychosocial symptoms.

In addition to the CBCL, the SDQ had a moderate to high correlation with measures of general and specific psychopathology. High correlations were found specifically for the Rutter scales, on which the SDQ is partly based (Goodman 1997), and for measures of depression and anxiety. This is contradictory to the low correlation found between the SDQ emotional and peer problems scales and the CBCL internalizing and social problems scales. However, as the SDQ correlated with specific measures of depression and anxiety here, the overlap between symptoms may have become greater and thus the correlations higher. Further, in community samples, SDQ scores also detect psychiatric diagnoses assigned by clinicians. Risk factors for developing psychosocial problems, such as poor health, seem to be associated with higher SDQ scores. This indicates that concurrent validity of the SDQ in comparison with different measures of psychopathology, psychiatric diagnoses, and risk factors is well established.

#### Capacity to Discriminate

The SDQ proves to be a good screening instrument, with high sensitivity and specificity for the total difficulties and impact scales. The percentage of children correctly identified by the SDQ as having a disorder is high, as is the percentage of children correctly identified by the SDQ as not having a disorder. A more detailed insight into the ability of the subscales to distinguish between community and clinical samples is reflected in the AUC values. Weighted AUC values indicate that, for teacher ratings only, the prosocial behavior and peer problems subscales distinguish between children with diagnoses, and those without, at the chance level. Prosocial behavior does not reflect child psychopathology, so it is not expected to distinguish between community and clinical samples. The peer problems scale again showed some inadequacy here.

However, we cannot infer from the sensitivity and specificity values which proportion of children with abnormal test results are truly abnormal (Altman and Bland 1994). When using the SDQ, we should therefore always consider the context, i.e., clinical versus community samples. If used in a community sample, quite a few children with clinical range SDQ results will actually be typically developing, i.e., false positives, due to low prevalence rates in the general population. In contrast, when the SDQ is used in a clinical sample, where prevalence rates are higher, fewer children will be false positives, but more will be false negatives. It is thus important to consider that the accuracy of the SDQ as a screening instrument varies

accordingly with the prevalence rates in a certain population. This underscores the need for using multiple diagnostic instruments in clinical or at risk settings, such as pediatric clinics.

#### Predictive Validity

Only three studies assessed the predictive validity using a longitudinal design. The results showed evidence of predictive validity, as SDQ scores predicted help seeking for psychosocial problems over a year. Two studies found evidence for SDQ scores predicting similar SDQ scores over a year. In addition, they clarified the role of prosocial behavior in the development of psychosocial problems. Prosocial behavior has not been found to be compatible with high levels of internalizing behavior and thus is not beneficial to children showing highly internalizing behaviors, which concurs with the literature (Hay 1994).

#### Conclusion

Overall, the 25-item SDQ shows strong psychometric properties. Shorter scales are usually less reliable compared to longer scales, which means they also tend to attenuate the validity (Streiner and Norman 1989). However, the SDQ's brevity did not substantially influence its psychometric properties. As for reliability, internal consistency of the total scales was satisfactory. Ratings showed sufficient reliability over time, and agreement between parents and teachers was relatively high. We should note here that these conclusions are stronger for teachers. Results concerning validity are less straightforward, but in general, we may state that the five-factor structure was confirmed by most studies, correlations with other measures of child psychopathology were high, and evidence for the screening ability of the SDQ was convincing. Predictive validity has not been studied extensively yet, so these findings should be interpreted with caution.

Additional attention should be directed to the necessity to conduct longitudinal studies that would examine the predictive validity of the SDQ and to the validation of the prosocial scale. Overall, the peer problems scale showed the weakest reliability and validity results that were most salient for parent ratings. The prosocial scale also showed some weaknesses, especially concerning internal consistency and capacity to discriminate. This notion should be familiar to researchers as these findings on the peer problems and prosocial behavior scales were extracted from previous studies. However, no interpretation of these findings has been proposed yet. A possible explanation of these findings lies in the concepts of prosocial behavior and peer problems.

In contrast to studies focusing on deviant behavior, studies assessing competence behaviors are relatively rare (Goodman 1994; Tremblay et al. 1992). As a consequence, the competence, or prosocial, construct has not been developed well in terms of what behaviors should be measured. A distinction in prosocial behavior is the Prosocial Orientation versus the Social Initiative dimension (Rydell et al. 1997). SDQ items are most comparable to the Prosocial Orientation dimension, which can be summarized as behaving smoothly in normal social interactions. In the Rydell et al., study, parent and teacher agreement was lower for the Prosocial Orientation than for the Social Initiative dimension. Possibly, the Social Initiative dimension consists of behaviors that are more easily observed (e.g., shy/hesitant with unfamiliar adults) than those of the Prosocial Orientation dimension (e.g., has ability to decode peers' feelings), and thus the comparable SDQ prosocial scale (e.g., considerate of other people's feelings). Behavior that is more difficult to observe may be more susceptible to inferences from raters, for example according to the relationship of the rater with the child (e.g., Ladd and Proffitt 1996). Inferences may be stronger for parents than for teachers in rating prosocial behavior, as internal consistency is lower for the former raters. This may be explained by the nature of the relationship with the child which differs clearly for parents versus teachers.

The peer problems scale showed low internal consistency values for both parent and teacher ratings. Peer problems are most often assessed via reports by children themselves (i.e., sociometrics) because children are regarded "insiders", whereas parents and teachers are regarded "outsiders" of the peer group. Judgments of peers are based on many and varied social interactions with those being assessed, which may be unknown to "outsiders" (Rubin et al. 2005). Assessment of peer problems by parents and teachers is further impeded by the adult perspective used to interpret children's social interactions, the relationship with the child and child's gender (Ladd and Mars 1986; Ladd and Proffitt 1996; Rubin and Coplan 1992). The outsider view combined with the mentioned rater biases may be responsible for the low internal consistency values for the peer problems scale found in our review.

Further, parents and teachers observe children in differing contexts, where different behaviors are shown. This may lead to lower values of internal consistency for both the peer problems and the prosocial subscale. As for rater bias, regardless of rater bias being a factor in the weak performance of subscales, it is important to be aware of rater bias when dealing with screening instruments. The application of screening instruments like the SDQ can be meaningful, in the sense that children are screened before psychosocial problems exacerbate, only if they are used appropriately.

Finally, it is important to note that results from this review are only applicable to the parent and teacher versions of the SDQ. The SDQ self-report version was not included in this review because it was not developed nor intended to be used for children younger than 12 years of age. From a developmental perspective, the use of traditional self-report questionnaires in children younger than 12 years of age has been questioned, and in children younger than 8 years discouraged. Due to limited linguistic, cognitive, and social-emotional abilities, children were not thought to provide reliable self-reports (Edelbrock et al. 1985; Fallon and Schwab-Stone 1994).

Recently, tests of using a puppet interview and computerized pictorial questionnaire have yielded results which point to promising psychometric results in children as young as 5–7 and 6–11 years (Measelle et al. 1998; Valla et al. 1994). However, the SDQ and the former interview methods differ greatly in respect to taking into account the developmental level of the elementary school child. The former interviews take into account the developmental level of the child by giving both visual (graphics) and auditory stimuli. The cognitive abilities of children below age 12 may not be sufficiently developed to adequately respond to the SDQ questions, which are presented only by visual verbal information (Edelbrock et al. 1985; Fallon and Schwab-Stone 1994). Therefore, we have focused on the parent and teacher versions of the SDQ in this review.

### Limitations

Some limitations of this review should be noted. First, the methodologies varied across the reviewed studies, making it sometimes impossible to extract data from those studies. Comparing these studies with each other was therefore difficult, and conducting a meta-analysis on the data was not possible. Second, many studies did not state which parent was used as a rater, making it hard to draw specific conclusions concerning rater bias. In addition, it was beyond our scope to consider rater psychopathology. Third, few studies were conducted using a longitudinal design, making it hard to draw robust conclusions regarding predictive validity. In addition, the reviewed studies did not give sufficient attention to validation of the prosocial scale. Future research should reveal whether the SDQ predicts psychosocial problems and whether the prosocial scale correlates with other measures of prosocial behavior.

### Implications

With these limitations in mind, the implications of these results for practice and research can be noted. This review offers researchers and clinicians a clear overview of the psychometric properties for the parent and teacher versions



of the SDQ for 4- to 12-year-olds. Reliability and validity results at the subscale level have been found weaker when compared to the results for the total scales. Therefore, caution is warranted when using and interpreting the subscales of the SDQ separately. Sanne et al. (2009) argued that the distinctiveness of the subscales is not convincing. An explanation for this may be the high comorbidity of psychosocial problems (Ford et al. 2003). Moreover, caution is warranted if a single informant reports on the SDQ, as results may not generalize to other contexts. The use of multiple informants should always be priority when using the SDQ. Most studies used parents and teachers, but possibilities of using other informants should be explored. For example, neighbors, daycare workers, or sports club coaches might be able to report on children's psychosocial problems. Future research should reveal whether these informants are able to assess psychosocial problems reliably.

For clinical practice in particular, the SDQ is a useful instrument for quickly assessing possible psychosocial problems. The results found in this review give rise to some specific implications at the subscale level.

First, the prosocial subscale shows some weaknesses in its psychometric properties, especially for the parent version. Low levels of prosocial behavior and high levels of aggression have been shown to increase the risk for future social adjustment difficulties (Coie et al. 1982; Crick 1996; Romano et al. 2005). Excessively high levels of prosocial behavior are also a risk factor for psychopathology (Hay 1994; Perren et al. 2007), underscoring the importance of assessing prosocial behavior. Therefore, when assessing prosocial behavior teacher ratings should always be included in addition to parent ratings. Further assessment of the child, for example by observing the child in the classroom or a naturalistic play situation, should reveal whether the reported lack of prosocial behavior is confirmed by a mental health specialist. When a child is referred for treatment, interventions target at the increase of prosocial behavior instead of the decrease of aversive behaviors (Coie and Koeppl 1990). This emphasizes the importance of assessing prosocial behavior adequately.

Second, the psychometric properties of the hyperactivity/inattention scale are adequate, and the SDQ should thus provide a reliable and valid report as to whether ADHD symptoms are present. However, when an ADHD diagnosis is suspected, identification of one of the subtypes Inattentive, Hyperactive-Impulsive, or Combined is required (American Psychiatric Association 1994). Further assessment may be done by using one of the many ADHD rating scales available, such as the SNAP-IV (Swanson 1992) or the SWAN (Swanson et al. 2001).

For the emotional symptoms subscale, psychometric properties are also adequate. However, in contrast to externalizing problems, internalizing problems are reported

more accurately by children themselves than by their parents and teachers (Edelbrock et al. 1985; Ederer 2004). Gaining insight into the child's subjective experience of its emotional symptoms is thus highly relevant and advisable in clinical settings.

The conduct problems subscale shows adequate reliability and validity. In order to assess whether a diagnosis of Oppositional Defiant Disorder or Conduct Disorder would be justified, additional assessment is indicated. Because children themselves tend to underestimate their externalizing problems, parents and teachers are particularly important in the further assessment of children presenting with conduct problems (Loeber et al. 1991).

Finally, the psychometric properties of the peer problems scale are quite weak in some respects. Assessing peer problems is complicated because children are considered as "insiders" who contribute unique information about their peer group. Possibly, it is difficult for parents and teachers to estimate the problems children experience in their peer group because they are "outsiders". Because of the difficulties with assessing peer problems, additional assessment is essential. The Perceived Competence Scale for Children (Harter 1982) is a very suitable measure for this purpose. Further, classroom observation is recommended (Wragg 1994).

The SDQ is not intended to be used as a psychiatric diagnostic instrument and therefore should not be utilized as such. As a screening instrument, the SDQ performs very well and adds to the field of early detection of child psychopathology. The SDQ has been translated into over 60 languages, which is a great benefit. However, norms are available only for six countries. Culture plays a role in the distribution and expression of psychosocial problems in society, and thus norms for every culture should be established. Results from studies assessing capacity to discriminate showed that the SDQ distinguishes well between children with and those without diagnoses. In populations at risk for psychosocial problems, such as children attending pediatric clinics, we recommend screening of all children referred to specialist services.

For research purposes, longitudinal designs should be employed in order to assess predictive validity more thoroughly. The SDQ is a promising instrument for researching developmental pathways, as it seems to be well validated, short, and acceptable. The teacher version shows strong psychometric properties, but our review shows that the parent version is at the focus of research (17 out of 48 studies studied only the parent version of the SDQ). However, researchers do not fully employ the use of a multiple informant approach. We do argue for such a multi-informant approach, as it is essential for children, their parents, and society when psychosocial problems are found at a young age.

**Acknowledgments** This work was granted by the Dutch Organization for Health Research and Care Innovation (ZonMW: 80-82435-98-8026-). The authors would like to thank Liesbeth Linssen for her advice regarding statistics.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

References marked with an asterisk indicate studies included in the review.

- Abikoff, H., Courtney, M., Pelham, W. E., & Koplewicz, H. S. (1993). Teachers' rating of disruptive behaviors: The influence of halo effects. *Journal of Abnormal Child Psychology*, *21*, 519–533.
- Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist 4–18 and 1991 profile*. Burlington, VT: University of Vermont Department of Psychiatry.
- Achenbach, T., McConaughy, S., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *101*, 213–232.
- Altman, D. G., & Bland, M. J. (1994). Diagnostic tests 2: Predictive values. *British Medical Journal*, *309*, 102.
- \*Alyahri, A., & Goodman, R. (2006). Validation of the Arabic Strengths and Difficulties Questionnaire and the development and well-being assessment. *Eastern Mediterranean Health Journal*, *12*(Suppl. 2), 138–146.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.
- \*Becker, A., Steinhausen, H. C., Baldusson, G., Dalsgaard, S., Lorenzo, M. J., Ralston, S. J., et al. (2006). Psychopathological screening of children with ADHD: Strengths and Difficulties Questionnaire in a pan-European study. *European Child and Adolescent Psychiatry*, *15*, 56–62.
- \*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), 11–16.
- Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (2002). *Cross-cultural psychology; Research and applications*. Cambridge, UK: Cambridge University Press.
- \*Bourdon, K. H., Goodman, R., Rae, D., Simpson, G., & Koretz, D. S. (2005). The Strengths and Difficulties Questionnaire: U.S. normative data and psychometric properties. *Journal of the American Academy of Child and Adolescent Psychiatry*, *44*, 557–564.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105.
- Canino, G., Shrout, P., Rubio-Stipec, M., Bird, H., Bravo, M., Ramirez, R., et al. (2004). The DSM-IV rates of child and adolescent disorders in Puerto Rico. *Archives of General Psychiatry*, *61*, 85–93.
- Carr, A. (2000). Family therapy: Concepts, process and practice. In A. Carr (Ed.), *The handbook of child and adolescent clinical psychology. A contextual approach* (2nd ed., pp. 93–94). London: Routledge.
- Caspi, A., Moffitt, T. E., Newman, D. L., & Silva, P. A. (1996). Behavioral observations at age 3 years predict adult psychiatric disorders: Longitudinal evidence from a birth cohort. *Archives of General Psychiatry*, *53*, 1033–1039.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (revised edition ed.). New York: Academic Press.
- Coie, J. D., Dodge, K. A., & Coppotelli, H. (1982). Dimensions and types of social status: A cross-age perspective. *Developmental Psychology*, *18*(4), 557–570.
- Coie, J. D., & Koeppel, G. K. (1990). Adapting intervention to the problems of aggressive and disruptive rejected children. In S. R. Asher & J. D. Coie (Eds.), *Peer rejection in childhood* (pp. 309–337). New York: Cambridge University Press.
- Corrigan, P. (2004). How stigma interferes with mental health care. *American Psychologist*, *59*(7), 614–625.
- Costello, E. J., Egger, H. L., & Angold, A. (2005). 10-Year research update review: The epidemiology of child and adolescent psychiatric disorders: I. Methods and public health burden. *Journal of the American Academy Child and Adolescent Psychiatry*, *44*, 972–986.
- Costello, E. J., Mustillo, S., Erkanli, A., Keeler, G., & Angold, A. (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of General Psychiatry*, *60*, 837–844.
- Crick, N. R. (1996). The role of overt aggression, relational aggression, and prosocial behavior in the prediction of children's future social adjustment. *Child Development*, *67*, 2317–2327.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *53*, 63–70.
- \*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*, 1159–1167.
- \*Du, Y., Kou, J., & Coghill, D. (2008). The validity, reliability and normative scores of the parent, teacher and self report versions of the Strengths and Difficulties Questionnaire in China. *Child and Adolescent Psychiatry and Mental Health*, *2*, <http://www.capmh.com/content/2/1/8>.
- Du, Y., Su, L., & Li, X. (1995). Usage of Conners Parent Symptom Questionnaire in boys with attention deficit hyperactivity disorder. *Chinese Journal of Clinical Psychology*, *1*, 44–45.
- DuPaul, G. J., Power, T. J., Anastopoulos, A. D., et al. (1998). *ADHD Rating Scale-IV: Checklists, norms, and clinical interpretation*. New York, NY: Guilford Press.
- Edelbrock, C., Costello, A. J., Dulcan, M. K., Kalas, R., & Calabro-Conover, N. (1985). Age differences in the reliability of the psychiatric interview of the child. *Child Development*, *56*, 265–275.
- Ederer, E. M. (2004). Mental health problems in young children: Self-reports and significant others as informants. *Psychology Science*, *46*, 123–140.
- \*Edmunds, S., Garratt, A., Haines, L., & Blair, M. (2005). Child Health Assessment at School Entry (CHASE) project: Evaluation in 10 London primary schools. *Child: Care, Health and Development*, *31*, 143–154.
- Egger, H. L., & Angold, A. (2006). Common emotional and behavioral disorders in preschool children: Presentation, nosology, and epidemiology. *Journal of Child Psychology and Psychiatry*, *47*, 313–337.
- Fallon, T., & Schwab-Stone, M. (1994). Determinants of reliability in psychiatric survey of children aged 6–12. *Journal of Child Psychology and Psychiatry*, *35*, 1391–1408.
- Field, A. P. (2001). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed- and random- effects methods. *Psychological Methods*, *6*(2), 161–180.

- Ford, T., Goodman, R., & Meltzer, H. (2003). The British child and adolescent mental health survey 1999: The prevalence of DSM-IV disorders. *Journal of the American Academy Child and Adolescent Psychiatry*, 42, 1203–1211.
- \*Glazebrook, C., Hollis, C., Heussler, H., Goodman, R., & Coates, L. (2002). Detecting emotional and behavioural problems in paediatric clinics. *Child: Care, Health and Development*, 29, 141–149.
- Goodman, R. (1994). A modified version of the Rutter parent questionnaire including items on children's strengths: A research note. *Journal of Child Psychology and Psychiatry*, 35, 1483–1494.
- \*Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38, 581–586.
- \*Goodman, R. (1999). The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. *Journal of Child Psychology and Psychiatry*, 40, 791–799.
- \*Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 1337–1345.
- \*Goodman, R., Ford, T., Corbin, T., & Meltzer, H. (2004). Using the Strengths and Difficulties Questionnaire (SDQ) multi-informant algorithm to screen looked-after children for psychiatric disorders. *European Child and Adolescent Psychiatry*, 13, 25–31.
- Goodman, R., Ford, T., Richards, H., Gatward, R., & Meltzer, H. (2000). The development and well-being assessment: Description and initial validation of an integrated assessment of child and adolescent psychopathology. *Journal of Child Psychology and Psychiatry*, 41, 645–655.
- \*Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000a). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *British Journal of Psychiatry*, 177, 534–539.
- \*Goodman, R., Iervolino, A. C., Collishaw, S., Pickles, A., & Maughan, B. (2007). Seemingly minor changes to a questionnaire can make a big difference to mean scores: A cautionary tale. *Social Psychiatry and Psychiatric Epidemiology*, 42, 322–327.
- \*Goodman, R., Renfrew, D., & Mullick, M. (2000b). Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *European Child and Adolescent Psychiatry*, 9, 129–134.
- \*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, 17–24.
- Gowers, S. G., Harrington, R. C., Whitton, A., Lelliott, P., Beevor, A., Wing, J., et al. (1999). Brief scale for measuring the outcomes of emotional and behavioural disorders in children. Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA). *British Journal of Psychiatry*, 174, 413–416.
- Harrington, R., Rutter, M., & Fombonne, E. (1996). Developmental pathways in depression: Multiple meanings, antecedents, and endpoints. *Development and Psychopathology*, 8, 601–616.
- Harter, S. (1982). The Perceived Competence Scale for children. *Child Development*, 53(1), 87–97.
- \*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, 644–651.
- Hay, D. F. (1994). Prosocial development. *Journal of Child Psychology and Psychiatry*, 35, 29–71.
- \*Hill, C. R., & Hughes, J. N. (2007). An examination of the convergent and discriminant validity of the Strengths and Difficulties Questionnaire. *School Psychology Quarterly*, 22, 380–406.
- Holland, D., & Dadds, M. (1995). *The diagnostic schedule for children, adolescents and parents*. Brisbane: Griffith University Press.
- \*Hysing, M., Elgen, I., Gillberg, C., Atle Lie, S., & Lundervold, A. J. (2007). Chronic physical illness and mental health in children. Results from a large-scale population study. *Journal of Child Psychology and Psychiatry*, 48, 785–792.
- Jackson, D. A., & King, A. R. (2004). Gender differences in the effects of oppositional behavior on teacher ratings of ADHD symptoms. *Journal of Abnormal Child Psychology*, 32(2), 215–224.
- \*Janssens, A., & Deboutte, D. (2009). Screening for psychopathology in child welfare: The Strengths and Difficulties Questionnaire (SDQ) compared with the Achenbach System of Empirically Based Assessment (ASEBA). *European Journal of Child and Adolescent Psychiatry*, 18, 691–700.
- Kataoka, S. H., Zang, L., & Wells, K. B. (2002). Unmet need for mental health care among U.S. children: Variation by ethnicity and insurance status. *American Journal of Psychiatry*, 159, 1548–1555.
- \*Kashala, E., Elgen, I., Sommerfelt, K., & Tylleskar, T. (2005). Teacher ratings of mental health among school children in Kinshasa, Democratic Republic of Congo. *European Child and Adolescent Psychiatry*, 14, 208–215.
- \*Klasen, H., Woerner, W., Wolke, D., Meyer, R., Overmeyer, S., Kaschnitz, W., et al. (2000). Comparing the German versions of the Strengths and Difficulties Questionnaire (SDQ-Deu) and the Child Behavior Checklist. *European Child and Adolescent Psychiatry*, 9, 217–276.
- \*Kaptein, S., Jansen, D. E. M. C., Vogels, A. G. C., & Reijneveld, S. A. (2008). Mental health problems in children with intellectual disability: Use of the Strengths and Difficulties Questionnaire. *Journal of Intellectual Disability Research*, 52, 125–131.
- \*Koskelainen, M., Sourander, A., & Kaljonen, A. (2001). The Strengths and Difficulties Questionnaire among Finnish school-aged children and adolescents. *European Child and Adolescent Psychiatry*, 9, 277–284.
- Kovacs, M. (1981). Rating scales to assess depression in school-aged children. *Acta Paedopsychiatrica*, 46, 305–315.
- Ladd, G. W., & Mars, K. T. (1986). Reliability and validity of preschoolers' perceptions of peer behavior. *Journal of Clinical Child Psychology*, 15, 16–25.
- Ladd, G. W., & Profilet, S. M. (1996). The Child Behavior Scale: A teacher report-measure of young children's aggressive, withdrawn and prosocial behaviors. *Developmental Psychology*, 32(6), 1008–1024.
- \*Lai, K. Y. C., Luk, E. S. L., Leung, P. W. L., Wong, A. S. Y., Law, L., & Ho, K. (2009). Validation of the Chinese version of the strengths and difficulties questionnaire in Hong Kong. *Social Psychiatry and Epidemiology*. doi:10.1007/s00127-009-0152-z.
- Loeber, R., Green, S. M., Lahey, B. B., & Stouthamer-Loeber, M. (1991). Differences and similarities between children, mothers and teachers as informants on disruptive child behavior. *Journal of Abnormal Child Psychology*, 19(1), 75–95.
- \*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, 357–363.
- \*Mathai, J., Anderson, P., & Bourne, A. (2002). The Strengths and Difficulties Questionnaire (SDQ) as a screening measure prior to admission to a child and adolescent mental health service (CAMHS). *Australian e-Journal for the Advancement of Mental Health*, 1, [www.auseinet.com/journal/vol1iss3/Mathai.pdf](http://www.auseinet.com/journal/vol1iss3/Mathai.pdf).
- \*Mathai, J., Anderson, P., & Bourne, A. (2004). Comparing psychiatric diagnoses generated by the Strengths and Difficulties

- Questionnaire with diagnoses made by clinicians. *Australian and New Zealand Journal of Psychiatry*, 38, 639–643.
- \*Matsuishi, T., Nagano, M., Araki, Y., Tanaka, Y., Iwasaki, M., Yamashita, Y., et al. (2008). Scale properties of the Japanese version of the Strengths and Difficulties Questionnaire (SDQ): A study of infant and school children in community samples. *Brain and Development*, 30, 410–415.
- Measelle, J. R., Ablow, J. C., Cowan, P. A., & Cowan, C. P. (1998). Assessing young children's views of their academic, social, and emotional lives: An evaluation of the self-perception scales of the Berkeley Puppet Interview. *Child Development*, 69, 1556–1576.
- \*Mellor, D. (2004). Furthering the use of the Strengths and Difficulties Questionnaire: Reliability with younger child respondents. *Psychological Assessment*, 16, 396–401.
- \*Mellor, D., & Stokes, M. (2007). The factor structure of the Strengths and Difficulties Questionnaire. *European Journal of Psychological Assessment*, 23, 105–112.
- Meltzer, H., Gatward, R., Goodman, R., & Ford, T. (2003). Mental health of children and adolescents in Great Britain. *International Review of Psychiatry*, 15, 185–187.
- \*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, 94–99.
- \*Muris, P., Meesters, C., & Van den Berg, F. (2003). The Strengths and Difficulties Questionnaire (SDQ): Further evidence for its reliability and validity in a community sample of Dutch children and adolescents. *European Child and Adolescent Psychiatry*, 12, 1–8.
- Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, 35, 250–256.
- \*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, 189–198.
- \*Parkes, J., White-Koning, M., Dickinson, H. O., Thyen, U., Arnaud, C., Beckung, E., et al. (2008). Psychological problems in children with cerebral palsy: A cross-sectional European study. *Journal of Child Psychology and Psychiatry*, 49, 405–413.
- \*Perren, S., Stadelmann, S., Von Wyl, A., & Von Klitzing, K. (2007). Pathways of behavioural and emotional symptoms in kindergarten children: What is the role of pro-social behaviour? *European Child and Adolescent Psychiatry*, 16, 209–214.
- Reynolds, C. R., & Richmond, B. O. (1978). What I think and feel: A revised measure of children's manifest anxiety. *Journal of Abnormal Child Psychology*, 6, 271–280.
- Rhee, S., Furlong, M. J., Turner, J. A., & Harari, I. (2001). Integrating strength-based perspectives in psychoeducational evaluations. *California School Psychologist*, 6, 5–17.
- Riley, A. W., Forrest, C. B., Starfield, B., Rebok, G. W., Robertson, J. A., & Green, B. F. (2004). The parent report form of the CHIP-Child Edition: Reliability and validity. *Medical Care*, 42, 210–220.
- Romano, E., Tremblay, R. E., Boulerice, B., & Swisher, R. (2005). Multilevel correlates of childhood physical aggression and prosocial behavior. *Journal of Abnormal Child Psychology*, 33(5), 565–578.
- \*Rothenberger, A., Becker, A., Erhart, M., Wille, N., Ravens-Sieberer, U., & the BELLA study group. (2008). Psychometric properties of the parent Strengths and Difficulties Questionnaire in the general population of German children and adolescents: Results of the BELLA study. *European Child and Adolescent Psychiatry*, 17(Suppl. 1), 99–105.
- Rubin, K. H., & Coplan, R. J. (1992). Peer relationships in childhood. In M. Bornstein & M. Lamb (Eds.), *Developmental psychology: An advanced textbook* (pp. 519–578). Hillsdale, NJ: Erlbaum.
- Rubin, K. H., Coplan, R. J., Chen, X., Buskirk, A., & Wojslawowicz, J. (2005). Peer relationships in childhood. In M. Bornstein & M. Lamb (Eds.), *Developmental science: An advanced textbook* (pp. 469–512). Mahwah, NJ: Erlbaum.
- Rutter, M. (1967). A children's behaviour questionnaire for completion by teachers: Preliminary findings. *Journal of Child Psychology and Psychiatry*, 8, 1–11.
- Rydell, M., Hagekull, B., & Bohlin, G. (1997). Measurement of two social competence aspects in middle childhood. *Developmental Psychology*, 33(5), 824–833.
- \*Samad, L., Hollis, C., Prince, M., & Goodman, R. (2005). Child and adolescent psychopathology in a developing country: Testing the validity of the Strengths and Difficulties Questionnaire (Urdu version). *International Journal of Methods in Psychiatric Research*, 13, 158–166.
- \*Sanne, B., Torsheim, T., Heiervang, E., & Stormark, K. M. (2009). The Strengths and Difficulties Questionnaire in the Bergen child study: A conceptually and methodically motivated structural analysis. *Psychological Assessment*, 21(3), 352–364.
- Scholte, E. M., & Van der Ploeg, J. D. (1998). *ADHD Vragenlijst (ADHD Questionnaire)*. Lisse: Swets & Zeitlinger.
- \*Sharp, C., Croudace, T. J., Goodyer, I. M., & Axtmann, D. (2005). The Strengths and Difficulties Questionnaire: Predictive validity of parent and teacher ratings for help-seeking behaviour over one year. *Educational and Child Psychology*, 22, 28–44.
- \*Shojaei, T., Wazana, A., Pitrou, I., & Kovess, V. (2008). The Strengths and Difficulties Questionnaire: Validation study in French school-aged children and cross-cultural comparisons. *Social Psychiatry and Psychiatric Epidemiology*. doi: 10.1007/s00127-008-0489-8.
- \*Smedje, H., Broman, J.-E., Hetta, J., & von Knorring, A.-L. (1999). Psychometric properties of a Swedish version of the "Strengths and Difficulties Questionnaire". *European Child and Adolescent Psychiatry*, 8, 63–70.
- Streiner, D. L., & Norman, G. R. (1989). *Health measurements scales*. Oxford, England: Oxford University Press.
- Swanson, J. M. (1992). *School-based assessments and interventions for ADD children*. Irvine, CA: KC.
- Swanson, J., Schuck, S., Mann, M., Carlson, C., Hartman, K., Sergeant, J., et al. (2001). *Categorical and dimensional definitions and evaluations of symptoms of ADHD: The SNAP and the SWAN Ratings Scales [Draft]* (2005) Available at: [http://www.adhd.net/SNAP\\_SWAN.pdf](http://www.adhd.net/SNAP_SWAN.pdf). Accessed May 28th, 2010.
- \*Syed, E. U., Hussein, S. A., Azam, S. I., & Khan, A. G. (2009). Comparison of Urdu version of Strengths and Difficulties Questionnaire (SDQ) and the Child Behavior Check List (CBCL) amongst primary school children in Karachi. *Journal of the College of Physicians and Surgeons Pakistan*, 9, 375–379.
- Tremblay, R. E., Vitaro, F., Gagnon, C., Piché, C., & Royer, N. (1992). A prosocial scale for the Preschool Behaviour Questionnaire: Concurrent and predictive correlates. *International Journal of Behavioral Development*, 15, 227–245.
- Valla, J. P., Bergeron, L., Bérubé, H., Gaudet, N., & St-Georges, M. (1994). A structured pictorial questionnaire to assess DSM-III-R-based diagnoses in children (6–11 years): Development, validity and reliability. *Journal of Abnormal Child Psychology*, 22, 403–423.
- \*Van Leeuwen, K., Meerschaert, T., Bosmans, G., De Medts, L., & Braet, C. (2006). The Strengths and Difficulties Questionnaire in a community sample of young children in Flanders. *European Journal of Psychological Assessment*, 22, 189–197.

- \*Van Roy, B., Veenstra, M., & Clench-Aas, J. (2008). Construct validity of the five-factor Strengths and Difficulties Questionnaire (SDQ) in pre-, early, and late adolescence. *Journal of Child Psychology and Psychiatry*, *49*, 1304–1312.
- \*Vogels, A. G. C., Crone, M. R., Hoekstra, F., & Reijneveld, S. A. (2009). Comparing three short questionnaires to detect psychosocial dysfunction among primary school children: a randomized method. *BMC Public Health*, *9*. doi:[10.1186/1471-2458-9-489](https://doi.org/10.1186/1471-2458-9-489).
- Weir, K., & Duveen, G. (1981). Further development and validation of the prosocial behavior questionnaire for use with teachers. *Journal of Child Psychology and Psychiatry*, *22*, 357–374.
- \*Widenfelt, B. M., van Goedhart, A. W., Treffers, P. D. A., & Goodman, R. (2003). Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *European Child and Adolescent Psychiatry*, *12*, 281–289.
- World Health Organization. (1992). *International classification of diseases 10th edition: Classification of mental and behavioural disorders*. Geneva: WHO.
- Wragg, E. C. (1994). *An introduction to classroom observation*. London: Routledge.
- \*Zwirs, B. W., Burger, H., Schulpen, T. W., & Buitelaar, J. K. (2008). Developing a brief cross-culturally validated screening tool for externalizing disorders in children. *Journal of the American Academy of Child and Adolescent Psychiatry*, *47*, 309–316.
- Zwirs, B. W., Burger, H., Schulpen, T. W., Wiznitzer, M., Fedder, H., & Buitelaar, J. K. (2007). Prevalence of psychiatric disorders among children of different ethnic origin. *Journal of Abnormal Child Psychology*, *35*, 556–566.