



Using the Dutch multi-informant Strengths and Difficulties Questionnaire (SDQ) to predict adolescent psychiatric diagnoses

J. Vugteveen¹ · A. De Bildt^{2,3} · C. A. Hartman⁴ · M. E. Timmerman¹

Received: 16 November 2017 / Accepted: 9 February 2018 / Published online: 24 February 2018
© The Author(s) 2018. This article is an open access publication

Abstract

Knowledge on the validity of the Strengths and Difficulties Questionnaire (SDQ) among adolescents is limited but essential for the interpretation of SDQ scores preceding the diagnostic process. This study assessed the predictive and discriminative value of adolescent- and parent-rated SDQ scores for psychiatric disorders, diagnosed by professionals in outpatient community clinics, in a sample of 2753 Dutch adolescents aged 12–17. Per disorder, the predictive accuracy of the SDQ scale that is contentwise related to that particular disorder and the SDQ impact scale was assessed. That is, 24 logistic regression analyses were performed, for each combination of DSM-IV diagnosis [4: Attention-Deficit/Hyperactivity Disorder (ADHD), Conduct/Oppositional Defiant Disorder (CD/ODD), Anxiety/Mood disorder, Autism Spectrum Disorder (ASD)], informant (3: adolescent, parent, both), and SDQ scale(s) (2; related scale only, related scale and impact scale). Additional logistic regression analyses were performed to assess the discriminative strength of the SDQ scales. The results show both fair predictive strength and fair discriminative strength for the adolescent- and parent-reported hyperactivity scales, the parent-reported conduct scale, and the parent-reported social and prosocial scales, indicating that these scales provide useful information about the presence of ADHD, CD/ODD, and ASD, respectively. The SDQ emotional scale showed to be insufficiently predictive. The findings suggest that parent-rated SDQ scores can be used to provide clinicians with a preliminary impression of the type of problems for ADHD, CD/ODD, and ASD, and adolescent for ADHD.

Keywords Adolescent self-report · Parent report · Mental health · Predicting psychiatric diagnoses

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00787-018-1127-y>) contains supplementary material, which is available to authorized users.

✉ J. Vugteveen
j.vugteveen@rug.nl

- ¹ Heymans Institute for Psychological Research, University of Groningen, Groningen, The Netherlands
- ² Department of Psychiatry, Child and Adolescent Psychiatry, University Medical Center Groningen (UMCG), Groningen, The Netherlands
- ³ Accare Child and Adolescent Psychiatry Groningen, Groningen, The Netherlands
- ⁴ Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion Regulation (ICPE), University of Groningen, University Medical Center Groningen (UMCG), Groningen, The Netherlands

Introduction

Adolescence is a developmental period associated with physical change, psychological development, and social adjustments while in the process of acquiring independence. The complexity of these coexisting processes leaves adolescents vulnerable to psychiatric disorders [1–3]. Estimates of the percentage of adolescents that are referred to outpatient clinics for youth mental or social health care vary between 10 and 15% [4, 5]. In outpatient clinics, screening questionnaires are often used as part of the diagnostic process by quickly generating a first impression of the problems at hand. Given the large numbers of adolescents and their parents that fill in such screening questionnaires, a continued research focus should be on how their scores can be helpful in the diagnostic process.

The Strengths and Difficulties Questionnaire (SDQ) is currently one of the most widely used screening instruments [6, 7]. The SDQ can be completed by adolescents themselves (aged 11–16) as well as by parents and/or teachers

(for children/adolescents aged 4–16). The questionnaire is relatively short and, as its name suggests, focusses on strengths (prosocial behaviour) as well as deficits (hyperactivity/inattention, conduct problems, emotional problems, and peer problems). In addition, the SDQ contains an impact scale which, if an adolescent experiences difficulties, can be used to indicate chronicity, distress, and social impairment for the adolescent as well as burden for others. The usefulness of the SDQ can be judged based on the principles associated with evidence-based assessment [8, 9]. The core idea of evidence-based assessment is to optimize individual assessment to suit the actual needs of the very individual. According to the principles of evidence-based assessment, an instrument can be a useful addition to a test battery if it is predictive of an important criterion [10]. The SDQ has repeatedly been evaluated from this perspective, considering different psychiatric disorders as the important criterion [11–15]. Because only a few of them have specifically focused on adolescents [13–15], more research on the accuracy of the SDQ for predicting diagnoses in adolescence is warranted. An important theme herein is that adolescence marks a shift towards using the adolescents themselves as informants, possibly combined with their parents, who are also used as informants during childhood, while increasingly less often using the teachers. At the same time, the parents' role as informants on their children's psychiatric problems slowly decreases and, for most types of problems, eventually ceases to exist.

In the two studies that we could trace in which a comparison was made between adolescent self-report and parent report [13, 15], SDQ scores were used to predict psychiatric disorders in any of three categories, namely, Attention-Deficit/Hyperactivity Disorder (ADHD), Conduct/Oppositional Defiant Disorder (CD/ODD) (both also referred to as externalizing disorders), and Anxiety/Mood disorder (also referred to as internalizing disorder). Each category of psychiatric disorders was predicted from the SDQ scale that is contentwise related to that particular category of disorders: the hyperactivity/inattention scale for ADHD, the conduct scale for CD/ODD, and the emotional scale for Anxiety/Mood disorder. In a large community sample, Goodman and colleagues [15] found that the parent is a better informant than the adolescent, both for externalizing and internalizing disorders. Adolescent self-report yielded low sensitivity rates for internalizing disorders (Anxiety/Mood disorder 0.28) and even lower sensitivity rates for externalizing disorders (ADHD 0.12, CD/ODD 0.15). Parent report yielded fair sensitivity rates for both externalizing disorders (ADHD 0.43, CD/ODD 0.40) and internalizing disorders (Anxiety/Mood disorder 0.39). Becker and colleagues compared adolescent self-report and parent report among adolescents in a clinical sample and also found the parent to be a better informant than the adolescent for both externalizing and

internalizing disorders, but the reliability of these findings is limited, because they were found in a rather small sample.

The current study contributes to knowledge about the construct validity of the SDQ by investigating how well diagnoses for specific psychiatric disorders can be predicted from self- or parent-reported SDQ scale scores in a large Dutch clinical sample of 2988 12–17-year-old adolescents referred to a mental health outpatient clinic. In line with earlier studies, we aim to predict ADHD, CD/ODD, and Anxiety/Mood disorder from (1) their contentwise-related scale (i.e., the hyperactivity/inattention scale, the conduct scale, and the emotional scale, respectively) and from (2) this contentwise-related scale combined with the impact scale. In addition, we explore how accurately Autism Spectrum Disorder (ASD) diagnoses can be predicted, considering the social scale and the prosocial scale as contentwise-related scales, as we presume that these scales could have some predictive value for ASD. We presume so because these scales are intended to provide a comprehensive first screening of social functioning. We acknowledge the existence and value of more specific and thorough ASD instruments, amongst others the Social Responsiveness Scale (SRS) [16] and the Children's Social Behaviour Questionnaire (CSBQ) [17] that contribute to charting the different aspects of ASD. However, such narrow-band instruments only measure ASD; they are different from broad-band screeners covering multiple types of psychopathology such as the SDQ. In line with the previous findings [13–15], we hypothesize that diagnoses for both externalizing disorders (i.e., ADHD and CD/ODD) and internalizing disorders (Anxiety/Mood) will be predicted fairly accurately. Based on findings from Goodman and colleagues [15] and general findings from psychopathology research among adolescents [18, 19], we hypothesize the parent to be a better informant than the adolescent for externalizing disorders. Concerning internalizing disorders, we hypothesize the adolescent to be the better informant, given that they have privileged access to their own less observable emotional difficulties like feeling persistent sadness. This hypothesis is in line with findings from general psychopathology research [18, 19], but deviates from findings by Goodman and colleagues [15] which suggest that the parent is the best informant for internalizing disorders too. As Goodman's findings were derived from a community sample instead of a clinical sample, as is the case in our current study, we base our hypothesis for internalizing disorders on general psychopathology literature. Regarding the prediction accuracy for ASD, we expect that parents are better informants than adolescents themselves. This hypothesis is based on the fact that self-report relies on the ability to recognize and verbalize emotions, intentions, and functioning, while the limitation in doing so is one of the core symptoms of ASD [20]. In addition, we expect higher levels of adolescent–parent agreement for the externalizing SDQ scales (i.e.,

hyperactivity/attention, conduct) than for the internalizing SDQ scale (i.e., emotional), as is consistent with findings in clinical samples using the Child Behaviour Checklist and Youth Self-report (CBCL and YSR, respectively) [21–23]. The SDQ impact scale is not exclusively related to any of the specific types of difficulties that are measured by the SDQ. To our knowledge, the prediction accuracy of the impact scale for specific types of disorders has not been investigated previously and we have no a priori expectations on its predictive strength. In addition to the predictive strength of each scale, we examine its discriminative strength by investigating how well each of the psychiatric disorders is predicted by their non-related scales.

To summarize, the aim of our study is twofold: (1) examine how well-specific types of psychiatric disorders, diagnosed in outpatient community clinics, can be predicted from SDQ scales in a large clinical sample and (2) investigate whether the accuracy of the prediction depends on the type of informant that was used.

Methods

Sample

Data were collected from adolescents who had been referred to one of the 29 outpatient clinics of an institution for child and adolescent psychiatry in the North of The Netherlands. The SDQ data were collected online during the intake assessment as part of routine outcome monitoring. The inclusion criteria for the sample were being a first time referral between January 1st of 2013 and December 31st 2015, falling within the age range of 12 through 17 and having received a clinical DSM-IV diagnosis. These criteria were met by 3826 adolescents. For 2988 (78.1%) of them, both the self-report and parent report SDQ data were available. Within this group, the mean age was 14.2 years (SD 1.6) among males (54.2%) and 14.6 years (SD 1.5) among females (45.8%).

Missing data

Of the total sample, 838 adolescents were missing SDQ data, from one SDQ informant (adolescent-reported SDQ data missing, $n = 148$; parent-reported SDQ data missing, $n = 291$), or both ($n = 399$). The scores from these adolescents were omitted from the analyses. Table 1 provides information about the age, sex, and diagnosed disorder distributions within the sample with missing SDQ data ($n = 838$) and within the study sample ($n = 2988$). The study sample was somewhat younger than the missing data sample [$t(3,826) = 9.20$, $p < 0.01$, 99% CI ($-0.45, -0.69$)]. Furthermore, in the study sample, ADHD diagnoses occurred

Table 1 Age, sex, and diagnosed disorder distributions within the study sample and the missing data sample

	Study sample	Missing data sample
<i>N</i>	2988	838
Mean age (SD)	14.4 (1.6)	14.9 (1.6)
Male/female	0.46/0.54 ^a	0.49/0.51
Diagnosed disorders ^b		
ADHD	0.29	0.21
CD/ODD	0.11	0.14
Anxiety/mood	0.40	0.46
ASD	0.21	0.23

ADHD Attention-Deficit/Hyperactivity Disorder, CD/ODD Conduct/ Oppositional Defiant Disorder, ASD Autism Spectrum Disorder

^aProportion of *N*

^bWithin both columns, the proportions related to the prevalence of the four disorders add up to more than 1 due to comorbidity of the disorders

relatively more frequently, and Anxiety/Mood disorders diagnoses less frequently, than in the missing data sample [ADHD: $z = 4.9$, $p < 0.01$, 99% CI (0.04, 0.13); Anxiety/Mood: $z = 3.5$, $p < 0.01$, 99% CI (0.02, 0.12)]. No evidence suggesting that the study sample differed from the missing data sample with respect to gender [male: $z = 1.3$, $p = 0.20$, 99% CI ($-0.03, 0.08$)] or the prevalence of CD/ODD and ASD [CD/ODD: $z = 2.6$, $p = 0.01$, 99% CI ($-0.01, 0.06$); ASD: $z = 1.4$, $p = 0.15$, 99% CI ($-0.02, 0.06$)] was found.

Strengths and Difficulties Questionnaire

Dutch translations of the parent and self-report versions of the SDQ were used [24]. The questionnaires consist of 33 items each. The first 25 items cover five scales, with four focusing on difficulties relating to behaviour, emotional functioning, hyperactivity/inattention, interaction with peers, and one focusing on the strength prosocial behaviour. The remainder of the items forms the impact scale which, if an adolescent has difficulties in one or more of the four difficulties scales, can be used to indicate chronicity, distress, social impairment and burden for others. The scales were computed in the standard manner [6, 7], resulting in scores ranging from 0 to 10 for each scale.

Clinical DSM-IV diagnosis

The adolescents' clinical diagnoses were established based on thorough diagnostic procedures by trained professionals in a multidisciplinary team, including at least a child- and adolescent psychiatrist, a child psychologist, and a specialized nurse. The diagnosis was based on information from various sources. In interviews with the adolescent, current functioning and complaints were assessed, and when

assumed relevant, standardized instruments were additionally administered, for example, the Anxiety Disorders Interview Schedule (ADIS) [25], or Autism Diagnostic Observation Schedule (ADOS) [26, 27]. Parents were interviewed separately from the adolescent about the developmental history of their child, and on current functioning and concerns. In addition, when assumed relevant, standardized instruments were administered, e.g., the ADIS-P [25] or Parent Interview for Child Symptoms (PICS) [28], when feasible the teacher(s) of the adolescent was (were) asked to provide information on daily functioning in school and on the adolescent's relationships with adults and peers with the Teacher Telephone Interview for ADHD and related disorders (TTI) [29].

The clinical diagnoses of the sample were grouped into the four DSM-IV categories: ADHD ($n = 872$, 29.2%), CD/ODD ($n = 323$, 10.8%), Anxiety/Mood disorder ($n = 1179$, 39.5%), and Pervasive Developmental Disorder (PDD; $n = 620$, 20.7%). In this study, we use the more current term ASD when referring to PDD. Per DSM-IV category, Table 2 provides information about comorbidity between these diagnoses.

Most notable is the frequent co-occurrence of CD/ODD with ADHD: of all adolescents with a CD/ODD diagnosis, 46.1% also received a diagnosis from the ADHD category. The other way around occurs much less frequently, as within the ADHD DSM-IV category 17.1% received a CD/ODD diagnosis.

Approximately, one out of six adolescents ($n = 506$, 16.9%) received a diagnosis that did not belong to any of these four categories; 'Eating disorder, not otherwise specified' ($n = 182$) or 'disorder of infancy, childhood or adolescence, not otherwise specified' ($n = 119$) were the most frequent.

Table 2 Prevalence of comorbidity per DSM-IV diagnosis category

Comorbid with					
DSM category	N^a	ADHD ^b	CD/ODD ^b	Anxiety/ Mood disorder ^b	ASD ^b
ADHD	872	–	0.17	0.14	0.14
CD/ODD	323	0.46	–	0.07	0.07
Anxiety/mood disorder	1179	0.10	0.02	–	0.10
ASD	620	0.20	0.04	0.19	–

ADHD Attention-Deficit/Hyperactivity Disorder, CD/ODD Conduct/ Oppositional Defiant Disorder, ASD Autism Spectrum Disorder

^aThe numbers in this column add up to more than 2988 (sample size) due to comorbidity

^bThe proportion of adolescents within each DSM category (row), also diagnosed with any of the other disorders

Statistical analyses

Per disorder, summary statistics (means and standard deviations) were calculated for all SDQ scales for both the self-report version and the parent version. Internal consistency information on the SDQ scales for both SDQ versions within in the study sample was retrieved by calculation of Cronbach's alpha coefficients. Per SDQ scale, the Cronbach's alpha coefficients of the two informants were compared with Feldt's test for dependent samples [30].

To assess potential informant effects in combination with disorder effects on SDQ scale scores, a repeated measures multivariate analysis of variance (rm-manova) with the SDQ scale scores as dependent variables and two within-subjects factors (informant and SDQ scale) was conducted for each of the four types of diagnosed disorders.

The strength of the informant agreement between the self-reported and the parent-reported scores per SDQ scale was examined through Pearson's correlations. Differences between correlation coefficients were tested using the Steiger's test.

The ability of the SDQ scales to predict a specific diagnosis was assessed via sensitivity and specificity rates using the 90th percentile score as cut-off score. In the absence of Dutch cut-off scores, we resorted to British population-based cutoffs [6].

The predictive value of the SDQ scales for the four disorders considered in this study was assessed by means of logistic regression analysis. These regression analyses were performed for each combination of disorder (4; ADHD, CD/ODD, Anxiety/Mood disorder, ASD) and informant (3; adolescent, parent, both). The predictive value of the SDQ scale contentwise related to the disorder involved was assessed (model 1: SDQ scale as a main effect), as well as the possible additive predictive value of the SDQ impact scale (model 2: SDQ scale and SDQ impact scale as main effects and interaction). This resulted in 24 analyses, all with the probability of receiving a particular disorder versus the probability of receiving any of the other disorders as the outcome.

To account for potential nonlinear relationships between predictor(s) and outcome, we considered the fit of two competing models for each predictor: first, a model containing the predictor as a linear effect and second, a model containing the predictor as a nonlinear effect via a restricted cubic spline with three knots (pp. 24–26) [31]. From these competing models, the model with the lowest value of Akaike's information criterion (AIC) was retained. The accuracy of the resulting prediction models was assessed with the area under the curve (AUC), corrected for optimism [32], thus expressing the so-called outsample prediction value. Using Harrell's guidelines, the optimism of the AUC values was estimated using 500 bootstrap samples [31]. In general, when AUC values are used to assess predictive strength,

values < 0.70 are considered ‘poor’, 0.70–0.80 ‘fair’, and ≥ 0.80 ‘good’ [33, 34].

We tested the model improvement resulting from the addition of the impact scale to the models including only the contentwise-related SDQ scales per disorder and the informant effect with DeLong’s method [35]. This method can be used to compare AUC values retrieved from the nested models (models 1 and 2 for predicting a particular disorder based on the same informant) and from correlated models (models 1 or models 2 for predicting a particular disorder based on different informants).

The discriminative strength of each SDQ scale was investigated by assessing how well each scale predicts the disorders it is contentwise unrelated to. The discriminative strength of a scale is considered fair when the AUC values indicating the prediction accuracy of the scale for all unrelated disorders is < 0.70, and poor when one or more AUC values ≥ 0.70.

For all statistical tests, a significance level of $\alpha = 0.01$ was used. All analyses were performed in the R version 3.2.3 [36]. The logistic regression analyses were performed using the rms package [37]. The comparisons of AUC values were performed using the pROC package [38].

Results

Summary statistics of SDQ scores

Table 3 presents internal consistency information for each of de SDQ scales for the adolescent self-reported and the parent-reported version.

Most internal consistency values (Cronbach’s alpha) for the SDQ scales range from 0.71 to 0.78 and are fairly similar across informants, exceptions being the conduct difficulties scale and the social difficulties scale. For these scales, the internal consistency values with the adolescent as informant are lower (0.59 and 0.55, respectively) than with the parent as informant (0.74 and 0.67, respectively).

Table 3 further presents means and standard deviations of SDQ scale scores for both the parent and the self-report version, per disorder and across all disorders, with the contentwise-related scale(s) per disorder printed in bold. Columnwise examination of Table 3 shows that the highest mean score per scale (and lowest for the prosocial scale which measures strengths) is found among the adolescents with the corresponding disorder (i.e., hyperactivity/inattention scale for ADHD; conduct scale for CD/ODD; emotional scale for Anxiety/Mood disorder; social scale and prosocial scale for ASD), as was expected. Note that a rowwise examination

Table 3 Per SDQ version (parent, adolescent) and per SDQ scale: descriptive statistics and internal consistency information

		SDQ scale							
		Difficulties					Strengths	Impact	
		Hyper	Conduct	Emotional	Social	Total	Prosocial	Impact	
		Parent report							
		<i>N</i>							
Entire sample	Cronbach’s alpha ^b	2988	0.76	0.74	0.75	0.67	0.78	0.75	0.71
	<i>M</i> (SD)		5.2 (2.8)	2.8 (2.4)	5.2 (2.8)	3.0 (2.3)	16.1 (6.4)	7.4 (2.2)	3.7 (2.6)
ADHD	<i>M</i> (SD)	872	7.1 (2.1)^a	3.6 (2.5)	4.2 (2.8)	2.6 (2.3)	17.5 (6.3)	7.3 (2.1)	3.6 (2.2)
CD/ODD	<i>M</i> (SD)	323	6.4 (2.4)	4.8 (2.4)	3.8 (2.6)	2.7 (2.2)	17.6 (6.6)	6.7 (2.1)	3.6 (2.4)
Anxiety/mood	<i>M</i> (SD)	1179	4.4 (2.6)	2.2 (2.1)	6.3 (2.4)	3.0 (2.2)	15.9 (6.2)	7.7 (2.1)	4.0 (2.7)
ASD	<i>M</i> (SD)	620	5.6 (2.6)	3.0 (2.5)	5.7 (2.7)	4.4 (2.3)	18.6 (6.3)	6.4 (2.4)	4.5 (2.5)
		Adolescent-report							
Entire sample	Cronbach’s alpha	2988	0.75	0.59	0.77	0.55	0.76	0.76	0.70
	<i>M</i> (SD)		5.3 (2.6)	2.5 (1.8)	4.5 (2.8)	2.3 (1.9)	14.7 (5.8)	7.9 (1.8)	2.3 (2.3)
ADHD	<i>M</i> (SD)	872	6.8 (2.1)	3.2 (1.9)	3.4 (2.4)	1.9 (1.7)	15.3 (5.3)	7.7 (1.7)	2.0 (2.0)
CD/ODD	<i>M</i> (SD)	323	5.7 (2.5)	3.7 (1.9)	2.7 (2.1)	2.0 (1.7)	14.1 (5.6)	7.5 (1.8)	1.8 (1.8)
Anxiety/Mood	<i>M</i> (SD)	1179	5.1 (2.4)	2.3 (1.7)	5.9 (2.5)	2.6 (1.9)	15.9 (5.8)	8.1 (1.7)	3.0 (2.5)
ASD	<i>M</i> (SD)	620	5.1 (2.6)	2.6 (1.9)	4.3 (2.7)	3.0 (2.1)	14.9 (6.1)	7.4 (2.0)	2.4 (2.4)

ADHD Attention-Deficit/Hyperactivity Disorder, CD/ODD Conduct/Oppositional Defiant Disorder, ASD Autism Spectrum Disorder, SDQ Strengths and Difficulties Questionnaire

^aFor each disorder the descriptives of the contentwise-related SDQ scale are presented in bold

^bPairwise comparisons of Cronbach alpha coefficients from both informants revealed differences in coefficients for the SDQ conduct ($p < 0.01$) and SDQ social ($p < 0.01$) scales, but not for the other scales ($p > 0.01$)

of the table is not very useful, because it only provides a comparison of mean scale scores within a group of adolescents with a particular disorder, thereby ignoring the fact that some types of behaviour are in general less prevalent among patients in outpatient clinics than others. In clinical practice, these differences between types of behaviour are corrected for through the use of cut-off values based on norms (i.e., scores that indicate the level of risk per range of SDQ scale scores) that differ across the SDQ scales.

Comparison of the mean parent and adolescent scores per scale provides an indication of the presence of a potential informant effect on the reported extent of problems. A few exceptions aside, parent-reported mean scores on the SDQ difficulties scales are higher than the equivalent adolescent-reported scores, indicating that parents report a greater degree of difficulties than adolescents. This also holds for the impact these difficulties have on daily life. In the same vein, adolescents report higher prosocial scores (SDQ strength scale) than parents for all disorders, indicating that adolescents are generally more positive about their strengths than their parents.

Both findings, i.e., (1) the highest (and for the prosocial scale the lowest) mean score per SDQ scale are found among the adolescents with the corresponding disorder and (2) parent-reported mean scores on the SDQ difficulties scales are generally higher than the equivalent adolescent-reported scores were associated with significant effects on all associated tests in the repeated measures MANOVA.

Informant agreement

Table 4 shows between-informant correlations per SDQ scale across the whole study sample.

The convergent correlations (correlation between adolescent and parent scores on the same SDQ scale; presented in bold) are positive and range from relatively weak (0.34 for impact) to moderately strong (0.58 for emotional). These values indicate limited agreement between adolescents and their parents. A comparison of informant agreement levels on each of the four SDQ difficulties scales revealed no significant differences between the scales, suggesting that adolescent–parent agreement does not depend on the type of problems the informants report on. Compared to informant agreement on the difficulties scales, significantly lower adolescent–parent agreement was found on the impact scale, suggesting that adolescents and parents more strongly agree on the existence of difficulties than on the impact of difficulties on the adolescents' life. Per SDQ strength or difficulty scale, the discriminant correlations (correlations between adolescent and parent scores on different SDQ scales) are significantly and substantially weaker than convergent correlations, which provides evidence for the discriminant validity of the SDQ scales.

In addition to Pearson correlations, we calculated convergent and discriminant intraclass correlation coefficients (see Online Resource). These coefficients show a similar pattern to the one described above.

Table 4 Between-informant (adolescent and parent) Pearson correlations per SDQ scale ($N = 2988$)

		Parent					Strengths	Impact	
		Difficulties					Prosocial	Impact	
SDQ scale		Hyper	Conduct	Emotional	Social	Total			
Adolescent	Difficulties	Hyper	0.54 ^{abd*}	0.23*	0.04	0.14*	0.35*	0.09*	0.24*
		Conduct	0.37*	0.54 *	0.01	0.15*	0.42*	−0.26*	0.17*
		Emotional	−0.18*	−0.18*	0.58 *	0.06	0.16*	−0.07*	0.15*
		Social	−0.02	< −0.01	0.26*	0.54 *	0.30*	−0.08*	0.23*
		Total	0.26*	0.19*	0.39*	0.31*	0.46 *	−0.10*	0.31*
	Strengths	Prosocial	−0.14*	−0.19*	0.01	−0.18*	−0.19*	0.41 *	−0.10*
	Impact	Impact	0.03	−0.03	0.35*	0.14*	0.20*	0.02	0.34 ^{c*}

*Correlation significant at the 0.01 level

^aCorrelations between adolescent and parent scores on the same SDQ scale (convergent correlations) are presented in bold

^bPairwise comparisons of the convergent correlations of the four SDQ difficulties scales (bold) revealed no significant differences between the four correlations. For all comparisons: $p > 0.01$

^cPairwise comparisons of convergent correlations on the SDQ impact scale and each of the SDQ difficulties scales showed that the convergent correlations on the impact scale are significantly lower than convergent correlations on each of the difficulties scales ($p < 0.01$)

^dPer SDQ strength or difficulties scale, pairwise comparisons of the convergent correlation (bold) with the discriminant correlations of the remaining strengths and difficulties scales revealed that the convergent correlations were significantly stronger than the discriminant correlations. For all comparisons: $p < 0.01$

Predicting disorders

Table 5 presents the sensitivity rate, specificity rate, and the diagnostic odds ratio for the contentwise-related SDQ scale(s) per type of disorder, using the 90th percentile in British population norms score as cut-off score.

A diagnostic odds ratio larger than 20 characterizes a useful test [39]. The diagnostic odds ratios in Table 5 range from 2.4 to 5.8, suggesting that the currently used cut-off values may not be appropriate for the clinical population at hand or the SDQ scales may not be useful predictors. To further investigate the value of the SDQ scales as predictors, a different approach that does not depend on cut-off values might be informative. Such an approach is assessing

the SDQ scales’ predictive strength through the estimation of prediction models.

Table 6 presents the estimated prediction accuracies of two prediction models per disorder, expressed in AUC values.

These values indicate how accurately the disorders can be predicted by either the contentwise-related scale (model 1) or the contentwise-related SDQ scale in combination with the SDQ impact scale (model 2).

The AUC values for the models containing only the contentwise-related SDQ scale per disorder (model 1) range from 0.63 (ASD, adolescent as single informant) to 0.80 (ADHD, both informants simultaneously), indicating poorly to fairly accurate predictions of the probability of receiving

Table 5 Sensitivity, specificity and the diagnostic odds ratio per SDQ version and disorder based on the British cut-off values [6]

Disorder ^a	N	SDQ scale	Informant					
			Adolescent			Parent		
			Sensitivity	Specificity	OR _D	Sensitivity	Specificity	OR _D
ADHD	872	Hyper	0.59	0.76	4.64 ^b	0.63	0.77	5.79
CD/ODD	323	Conduct	0.32	0.87	3.24	0.67	0.72	5.14
Anxiety/mood	1179	Emotional	0.46	0.84	4.64	0.76	0.52	3.52
ASD	620	Social	0.12	0.95	2.41	0.63	0.69	3.82
		Prosocial	0.09	0.97	2.97	0.23	0.91	3.01

ADHD Attention-Deficit/Hyperactivity Disorder, CD/ODD Conduct/Oppositional Defiant Disorder, ASD Autism Spectrum Disorder, SDQ Strengths and Difficulties Questionnaire

^aFor each disorder the descriptives of the contentwise-related SDQ scale is presented

^bThe diagnostic odds ratio $OR_D = (sensitivity \times specificity) / ((1 - sensitivity) \times (1 - specificity))$

Table 6 AUC values (corrected for optimism) for models 1 and 2 per disorder

		Informant			Comparing informants		
		A	P	B	AP	AB	PB
ADHD (n = 872)							
Model 1	Hyper	0.74	0.78	0.80	*	*	*
Model 2	Incl impact	0.77	0.80	0.82	*	*	*
CD/ODD (n = 323)							
Model 1	Conduct	0.69	0.76	0.77	*	*	ns
Model 2	Incl impact	0.76	0.78	0.81	ns	*	*
Anxiety/mood disorder (n = 1179)							
Model 1	Emotional	0.73	0.69	0.74	*	*	*
Model 2	Incl impact	0.73	0.70	0.75	*	*	*
ASD (n = 620)							
Model 1	Social + prosocial	0.63	0.74	0.74	*	*	ns
Model 2	Incl impact	0.64	0.74	0.74	*	*	ns

Due to the large sample size used in the analyses, the presented optimism-corrected values are equal to the raw AUC values, with the exception of (1) ASD model 1 with the adolescent as informant (raw AUC 0.64) and (2) ASD model 2 with both informants (raw AUC 0.75)

A adolescent, P parent, B both adolescents and parents, ADHD Attention-Deficit/Hyperactivity Disorder, CD/ODD Conduct/Oppositional Defiant Disorder, ASD Autism Spectrum Disorder

*Difference between informants significant at the 0.01 level, ns not significant

a certain diagnosis. Table 6 shows the highest AUC values for ADHD and the lowest for CD/ODD and ASD when the adolescent is used as a single informant and Anxiety/Mood disorder when the parent is the single informant.

Extending the models with the main effect of the impact scale and its interaction with the contentwise-related scale (model 2) improves the accuracy of the prediction for ADHD and CD/ODD (average AUC improvement of 0.02 and 0.04 across informants, respectively). For both informants separately and for both informants combined, the change in AUC values is statistically significant at an $\alpha = 0.01$ level. For Anxiety/Mood disorder and ASD, prediction accuracy does not improve when the impact scale is added to the models.

Informant effects per disorder

To assess potential informant effects, a comparison per model (i.e., models 1 and 2) was made between the predictive values (see AUC values and statistical tests in Table 6) of the models based on only adolescent information, the models based on only parent information and the models based on both adolescent and parent information.

Attention-deficit/hyperactivity disorder The parent is the best single informant when either model 1 or model 2 is used for the prediction of ADHD. Compared to using either single informant, the prediction accuracy of the models slightly improves when both informants are used simultaneously.

Conduct/oppositional defiant disorder The parent is the best single informant when model 1 is used to predict CD/ODD, and using both informants does not improve the prediction accuracy. The AUC values for model 2 do not identify either one of the informants to be superior over the other. Using the informants simultaneously leads to a slight increase in prediction accuracy of model 1 when compared to using the adolescent as informant, but not compared to the parent as single informant. For model 2, the combination of both informants is superior to using either single informant.

Anxiety/mood disorder The adolescent is the best single informant, both when model 1 and when model 2 is used to predict Anxiety/Mood disorder. Using both informants simultaneously hardly improves the prediction accuracy of models 1 and 2, but the improvement is significant.

Autism spectrum disorder The parent is the best single informant for the prediction of ASD for both models. Adding the information provided by the adolescent does not seem to improve the accuracy of the predictions based on parent information.

Discriminative strength

Table 7 presents how well each disorders is predicted by each of the SDQ scales. The discriminative strength of each SDQ scale can be assessed by examining how well each disorder is predicted by their contentwise-unrelated scales.

The SDQ hyperactivity scale, conduct scale, social scale, and prosocial scale each poorly predict the disorders they are not intended to predict well, regardless of the informant that was used. These findings indicate fair discriminative strength for each of these four scales. The SDQ emotional scale poorly predicts the disorders that it was not intended to predict when the parent is used as informant, and fairly with the adolescent as informant. This indicates fair

Table 7 AUC values (corrected for optimism) for each SDQ scale per disorder

SDQ scale	Informant		
	A	P	B
ADHD ($n = 872$)			
Hyper	0.74^a	0.78	0.80
Conduct	0.64	0.64	0.65*
Emotional	0.67	0.64	0.68
Social	0.58	0.57	0.59
Prosocial	0.55	0.55	0.56*
CD/ODD ($n = 323$)			
Hyper	0.54*	0.64	0.64
Conduct	0.69	0.76	0.77
Emotional	0.72	0.66	0.72
Social	0.55	0.54	0.55
Prosocial	0.57	0.61	0.61*
Anxiety/mood disorder ($n = 1179$)			
Hyper	0.55*	0.63	0.65
Conduct	0.56	0.62	0.62
Emotional	0.73	0.69	0.74
Social	0.57	0.53	0.58
Prosocial	0.55	0.57	0.57*
ASD ($n = 620$)			
Hyper	0.53	0.54	0.59
Conduct	0.49*	0.54	0.54
Emotional	0.53	0.56	0.60
Social	0.61	0.72	0.72
Prosocial	0.59	0.66	0.66

ADHD Attention-Deficit/Hyperactivity Disorder, CD/ODD Conduct/ Oppositional Defiant Disorder, ASD Autism Spectrum Disorder, SDQ Strengths and Difficulties Questionnaire, A adolescent, P parent, B both adolescents and parents

*The raw AUC values are 0.01 larger than the presented optimism-corrected values

^aFor each disorder the descriptives of the contentwise-related SDQ scale are presented in bold

discriminative strength with the parent and poor discriminative strength with the adolescent as informant.

Discussion

The aim of our study was to examine how well-specific types of psychiatric disorders, diagnosed in outpatient community clinics, could be predicted from Dutch SDQ scales in a large clinical sample of 12–17 years and to investigate whether the accuracy of the prediction depended on the type of informant that was used. Cut-off values are not available for Dutch adolescents. Using the 90th percentile in British population norms scores [6] as cut-off scores we found sensitivity rates, specificity rates and diagnostic odds ratios that suggested that either the used cut-off values were not appropriate for the clinical population at hand or that the SDQ scales were not useful as predictors for the disorders (ADHD, CD/ODD, Anxiety/Mood disorder, ASD). In the absence of any further indication of appropriate cut-off scores for the Dutch population and knowing that working with cut-off values entails using limited information from SDQ scale scores (as they are divided into 3–4 categories only), we proceeded to investigate the predictive and discriminative strength of the SDQ scales by estimating prediction models. For each SDQ scale (hyperactivity, conduct, emotional, social, and prosocial) and per informant (adolescent, parent, or both), prediction models were used to investigate the scale's predictive and discriminative strength. A scale's predictive strength was examined by assessing how well the scale predicted the disorder it was contentwise related to. The discriminative strength of each scale was investigated by assessment of how well the scale predicted the disorders it was contentwise unrelated to. As was hypothesized, we found that diagnoses for externalizing disorders (i.e., ADHD and CD/ODD) and internalizing disorders (Anxiety/Mood) could be predicted fairly accurately from their contentwise-related SDQ scale(s), which are the SDQ hyperactivity/inattention scale, conduct scale and emotional scale for ADHD, CD/ODD, and Anxiety/Mood disorder, respectively. We further found the parent to be the best informant for externalizing disorders, whereas the adolescent was the best informant for internalizing disorders, as is consistent with our hypothesis that was based on general findings from psychopathology research among adolescents [18, 19]. Our findings indicate fair predictive strength for the SDQ hyperactivity scale regardless of the informant that was used. Furthermore, the findings show fair predictive strength for the conduct scale with the parent as informant and the emotional scale with the adolescent as informant. Similar levels of adolescent–parent agreement were found across the difficulties scales, which is in contrast with our hypothesis on higher levels of agreement for the externalizing SDQ scales (i.e., hyperactivity/

attention, conduct) compared to the internalizing SDQ scale (i.e., emotional). A possible explanation for this deviation is that the group of adolescents with a diagnosis for Anxiety/Mood disorder in our sample consists of relatively many adolescents with anxiety problems (59.5%), few with mood problems (26.4%), and some with both (14.2%). Previous research suggests that, although both regarded as internalizing disorders, anxiety is more easily observable than mood problems [40]. Anxiety might, therefore, not only be relatively accurately reported by the adolescent but also by the parent, resulting in a higher level of adolescent–parent agreement. Regarding the possible additional value of including the impact scale, we did not state a hypothesis. We found that prediction accuracy for only ADHD and CD/ODD disorders improved when the impact of problems was included in the prediction models. This suggests that the impact scale contributes to the prediction of externalizing but not internalizing disorders among a clinical population.

Compared to other studies that assessed the SDQ's predictive abilities among adolescents, our study is the first in its attempt to predict ASD from the SDQ. It remains unclear why Goodman [15], He [14], Becker [13], and their respective colleagues refrained from doing so in their studies among adolescents, but in another study (involving children and adolescents without distinguishing between the two) Goodman offers an explanation for omitting patients with ASD: “First, the SDQ is clearly focused on common forms of psychopathology and does not include the sorts of questions that would allow the recognition of autistic or psychotic disorders with confidence. Second, it is generally easy to recognize children at risk of psychosis or autism from the referral letter, so there would be little additional merit in predicting these disorders from prior SDQs even if this were possible. Third, new referrals with these disorders are relatively rare in district clinics...” (p. 130) [41]. We only partially agree with Goodman. ASD is a relatively common disorder, with an estimated prevalence up to 1.5% in the general community [42]. In our study, no less than 20.7% of the total sample had received an ASD diagnosis. However, characteristics of adolescents referred to outpatient clinics may differ from adolescents in district clinics, which was the setting of Goodman's study. The adolescents in the current sample seem to have managed to function well enough to avoid an earlier referral, suggesting that the adolescents with ASD in our sample were relatively high-functioning. Our sample is, therefore, not fully representative of the population of adolescents with ASD. Although there is no SDQ scale that is specifically designed to measure autistic behaviour, which was mentioned by Goodman and colleagues as one of the reasons not to include ASD in their study, ASD is defined by social problems. Our findings suggest that, with the parent as informant, ASD can be fairly accurately predicted

from the SDQ social and prosocial scales, indicating fair predictive strength of these SDQ scales combined. Thus, we conclude that for high-functioning adolescents with ASD, parent-rated social difficulties and prosocial behaviour can serve as a fairly accurate first-impression proxy of the potential presence of ASD.

To be useful for assessment purposes, the SDQ scales should not only be predictive of the disorder they are contentwise related to, but they should also be able to discriminate between disorders. All but one SDQ scale showed fair discriminative strength, regardless of the informant that was used. The exception was the emotional scale. Although discriminating fairly well with the parent as informant, the emotional scale did not with the adolescent as informant. Based on the adolescent-reported emotional scale CD/ODD was unintendedly predicted fairly accurate. These findings could indicate that the SDQ emotional scale with the adolescent as informant is of limited use. However, it might be that Anxiety/Mood disorders are underdiagnosed among adolescents with CD/ODD in the sample used in this study. Literature suggests the rates of comorbid Anxiety/Mood disorders among youth with CD/ODD disorders are approximately 40% [43], whereas this specific type of comorbidity was only found in 7% of adolescents with CD/ODD in the sample under study here. If CD/ODD would be indeed underdiagnosed in the current sample, this provides an alternative explanation for the predictive value of the adolescent-reported emotional scale for CD/ODD. The parent-reported emotional scale does not appear to be predictive for CD/ODD, possibly because the parent showed to be a poorer informant for emotional problems.

Considering SDQ scales that show both fair predictive strength and fair discriminative strength as useful scales for providing clinicians with a preliminary impression of the type of problems at hand, we conclude that the SDQ hyperactivity scale is useful for providing information about the potential presence of ADHD, regardless of the informant that was used. The SDQ conduct scale and the combination of the SDQ social and prosocial scales are useful for indicating the presence of CD/ODD and ASD, respectively, with the parent as informant. With the adolescent as informant, these scales' predictive strength is inadequate. Furthermore, the SDQ emotional scale is not useful for assessment as it is not sufficiently discriminative with the adolescent as informant and not sufficiently predictive with the parent as informant; the combination of the adolescent and the parent as informants, does not provide a solution.

Consistent with the previous research, we investigated informant agreement through the calculation of correlations between adolescent and parent scores per SDQ scale. We found similar levels of adolescent–parent agreement for externalizing and internalizing difficulty scales. This finding deviates from our hypothesis, which was based on earlier

findings that adolescent–parent correlations were higher for externalizing scales than for internalizing scales [22, 23]. For various reasons, many studies do not proceed after investigating informant agreement. We strongly recommend to additionally study the association between both adolescent and parent-rated scores and the diagnosis which the adolescents received, because without it, informant agreement is limitedly useful as it does not provide information about which, if any, of the informants is a good informant. To that end, we performed logistic regression analyses, which identified a best informant for each disorder, with the exception of anxiety/mood disorders.

The findings of the current study emphasize the need for an assessment method that combines scores from SDQ difficulties and strength scales with the SDQ impact scale and, as literature on evidence-based assessment suggests too [8], combines information provided by multiple informants. To be optimally useful in clinical practice, this method should result in a probability prediction per type of disorder for each individual. In our view, the methods that are currently most widely used in clinical practice do not fully suffice. That is, using cut-off values results in a categorization into one of three or four (depending on the cut-off solution used) categories per person per SDQ scale. It does not allow combining information from multiple SDQ scales or informants. The alternative is utilizing the algorithm proposed by Goodman and colleagues [41], which combines SDQ difficulties scales with the impact scale and combines information from informants and results in a blunt 'unlikely', 'possible', or 'probable' rating per person per disorder (emotional, conduct, or hyperactivity disorder). This method requires information from all informants (adolescent, parent, and teacher), which limits its applicability in clinical practice. A useful alternative, would be to use a nomogram [44] derived from a prediction model, estimated based on both community samples and clinical samples. A nomogram is a visual tool that allows the clinical user to retrieve an individual's probability of receiving a particular diagnosis. This tool also visualizes effect sizes per predictor and how predictors interact with each other in predicting the probability of receiving one of the types of disorders.

Strengths and limitations

Our study focuses on the validity of the SDQ within a clinical setting. Our clinical sample is large, compared to clinical samples of adolescents from other studies, and the sample size per disorder is considerable. In that respect, our study clearly surpasses previous studies. Note that our findings pertain to a clinical population and hence do not allow us to infer that the SDQ is useful for detection of psychosocial problems in the general population.

The clinical diagnoses that were predicted in this study, were established by a multidisciplinary team of trained professionals, based on thorough diagnostic procedures. During these procedures, information was gathered from the adolescents, their parent(s) and, if deemed necessary, their teacher. We realize that this process was not compliant with the STandardized Reporting of Diagnostic assessment guidelines (STARD) [45, 46], because the diagnoses were only partially corroborated with standardized diagnostic instruments and can thus not be regarded as standardized diagnoses. Besides, the literature shows limited agreement between clinician-generated diagnoses and diagnoses generated from standardized procedures [47, 48], indicating that the reliability of diagnoses used in this study is potentially limited. The clinician-generated diagnoses in this study were established by extensively trained and experienced professionals. As a result, these diagnoses can be regarded as ‘true’ in the sense that these were the actual diagnoses that elicited a certain type of treatment. While the use of instruments in a fully standardized procedure, an approach frequently employed in scientific studies, is presumably more reliable, it does not fully represent clinical practice. Given the fact that the clinical diagnoses that were used in the current study are not beyond any doubt, we feel inclined to advocate some cautiousness interpreting the results of our study.

The SDQ data were collected at the start of the diagnostic process as part of the Routine Outcome Measurement (ROM). The ROM data are primarily collected for insurance and policy-making purposes. These data were accessible to the multidisciplinary team during their assessment of the adolescents functioning, which is in conflict with the aforementioned STARD guidelines, but typically, the data are not used for diagnostic considerations. This is actually one of the main reasons why we conducted the current study, i.e., given that adolescents and their parents spend time filling in this questionnaire, we wanted to provide a thorough evaluation whether and, if so, how this information can be put to use for their own benefit, given that currently it tends to be completely ignored. Hence, though it cannot be ruled out that the ROM data might have influenced the outcome of the diagnostic process for some adolescents, we expect the actual influence of the SDQ scores on the clinical diagnosis to be negligible.

Both limitations just discussed (i.e., the absence of a fully standardized assessment procedure and the accessibility of SDQ information during assessment) might have had an effect on the predictive value of the SDQ scales. The potential effects are in opposite direction. First, the use of clinically generated diagnoses may have tempered the effects that were found in this study, because a more reliable outcome measure could potentially have been more accurately predicted. Second, the accessibility of the SDQ information during the health-care professional’s assessment may have

affected some of the diagnoses assigned by the professionals, consequently leading to overestimation of the SDQ scales’ predictive abilities. As we have no way to estimate the size of these effects, we do not know their net direction and size.

In our study, we took comorbidity of disorders into account by allowing multiple diagnoses per adolescent. We performed the analyses per type of disorder. Further research is needed to investigate if combinations of SDQ scales can be used to predict specific types of comorbidity. In addition, it could be informative to further consider the heterogeneity within a group with a specific disorder. As far as we could trace, all the previous studies that assess the SDQ’s predictive validity—including ours—investigated how well disorders can be predicted from one or more SDQ scales. By doing so, we neglect the fact that most adolescents with, for instance, an Anxiety/Mood disorder score relatively high on the SDQ emotional difficulties scale, but not all of them score equally high or low on the other SDQ scales. For clinical practice, it could be highly useful to identify SDQ score profiles and investigate how well these profiles predict types of diagnosis. In other words, the next step would be to take diversity in SDQ scores as a starting point and then predict diagnoses as opposed to examining what adolescents with a specific diagnosis have in common, as has been done so far. Such profile information can, as was suggested before, be used to estimate an individual’s probability at each of the four types of diagnoses [44].

Implications

Clinical assessment is aimed at diagnosing and planning treatment. It is important that the outcome of clinical assessment is accurate, because the stakes are high for individuals in need of care. Therefore, it is important that assessment is thorough and that only useful tools are used. Considering the SDQ as such a potentially useful tool, we investigated the extent to which Dutch SDQ scales can be used to predict diagnoses and how well they discriminate between different types of diagnoses. The results of this study show that for adolescents referred to an out-patient clinic the SDQ hyperactivity scale is useful for providing information about the potential presence of ADHD, regardless of the informant that was used. The parent-reported SDQ conduct scale and the combination of the parent-reported SDQ social and prosocial scales are informative about the presence of CD/ODD and ASD, respectively. The SDQ-emotional scale is insufficiently indicative of the presence of Anxiety/Mood disorders, regardless of the informant that was used. It is important to notice that even the most accurate predictions based on the SDQ scales are far from perfect and cannot replace thorough clinical assessment. In addition, we caution that it is not informative to compare SDQ scale scores within a single individual to gain insight into their relative

problem levels, because some types of behaviour are generally less prevalent or less common than the others. This holds for the general population as well as for (specific) outpatient populations. That, for example, makes a scale score of six (relatively high) on the conduct scale incomparable to a scale score of six on the hyperactivity/attention scale (only moderately high). Cut-off values or normed scores, based on either the general population or children with a specific disorder, may be used for cross-disorder comparisons. The results of this study suggest that it is useful for clinicians to take the SDQ scales, except the SDQ emotional scale, into account as a first step in the diagnostic process to possibly steer attention towards one or more specific types of disorders, which should then be more thoroughly considered by clinicians. The parent showed to be a useful informant for ADHD, CD/ODD, and ASD, and the adolescent for ADHD. For clinical practice, in which it is often challenging to get both the adolescent and the parent to fill in a questionnaire, these findings suggest that it is most useful to ask the parent to fill in the SDQ.

Funding This research was funded by The Netherlands Organization for Health Research and Development (ZonMw, nr. 729300105).

Compliance with ethical standards

Approval This study was approved by the ethics committee of the Heymans Institute for Psychological Research of the University of Groningen in the Netherlands.

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Abela JR, Hankin BL (2008) Handbook of depression in children and adolescents. Guilford Press, New York
- Olfson M, Blanco C, Wang S, Laje G, Correll CU (2014) National trends in the mental health care of children, adolescents, and adults by office-based physicians. *JAMA Psychiatry*. 71:81–90
- Balazs J, Miklósi M, Keresztény Á, Hoven CW, Carli V, Waserman C, Apter A, Bobes J, Brunner R, Cosman D (2013) Adolescent subthreshold-depression and anxiety: psychopathology, functional impairment and increased suicide risk. *J Child Psychol Psychiatry*. 54:670–677
- Olfson M, Druss BG, Marcus SC (2015) Trends in mental health care among children and adolescents. *N Engl J Med* 372:2029–2038
- Reijneveld SA, Wieggersma PA, Ormel J, Verhulst FC, Vollebergh WA, Jansen DE (2014) Adolescents' use of care for behavioral and emotional problems: types, trends, and determinants. *PLoS One* 9(4):e93526. <https://doi.org/10.1371/journal.pone.0093526>
- Goodman R (1997) The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol 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. *J Child Psychol Psychiatry* 40:791–799
- Hunsley J, Mash EJ (2007) Evidence-based assessment. *Annu Rev Clin Psychol* 3:29–51
- Youngstrom E (2013) Future directions in psychological assessment: combining evidence-based medicine innovations with psychology's historical strengths to enhance utility. *J Clin Child Adolesc Psychol* 42:139–159
- Youngstrom E, Frazier T (2013) Evidence-based strategies for the assessment of children and adolescents: measuring prediction, prescription, and process. In: Youngstrom EA, Van Meter A, Frazier TW, Hunsley J, Prinstein MJ, Ong M, Youngstrom JK (eds) Evidence-based assessment as an integrative model for applying psychological science to guide the voyage of treatment. Clinical psychology: science and practice. University of North Carolina, Chapel Hill
- Brøndbo PH, Mathiassen B, Martinussen M, Heiervang E, Eriksen M, Moe TF, Sæther G, Kvernmo S (2011) The Strengths and Difficulties Questionnaire as a screening instrument for norwegian child and adolescent mental health services, application of UK scoring algorithms. *Child Adolesc Ment Health* 5:32
- Klasen H, Woerner W, Wolke D, Meyer R, Overmeyer S, Kaschnitz W, Rothenberger A, Goodman R (2000) Comparing the German versions of the Strengths and Difficulties Questionnaire (SDQ-Deu) and the child behavior checklist. *Eur Child Adolesc Psychiatry* 9:271–276
- Becker A, Hagenberg N, Roessner V, Woerner W, Rothenberger A (2004) Evaluation of the self-reported SDQ in a clinical setting: Do self-reports tell us more than ratings by adult informants? *Eur Child Adolesc Psychiatry* 13:17–24
- He J, Burstein M, Schmitz A, Merikangas KR (2013) The Strengths and Difficulties Questionnaire (SDQ): the factor structure and scale validation in US adolescents. *J Abnorm Child Psychol* 41:583–595
- Goodman R, Ford T, Simmons H, Gatward R, Meltzer H (2000) Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *Br J Psychiatry* 177:534–539
- Constantino JN, Gruber CP (2005) Social responsiveness scale (SRS). Western Psychological Services, Los Angeles
- Hartman CA, Luteijn E, Serra M, Minderaa R (2006) Refinement of the Children's Social Behavior Questionnaire (CSBQ): an instrument that describes the diverse problems seen in milder forms of PDD. *J Autism Dev Disord* 36:325–342
- Cantwell DP, Lewinsohn PM, Rohde P, Seeley JR (1997) Correspondence between adolescent report and parent report of psychiatric diagnostic data. *J Am Acad Child Adolesc Psychiatry* 36:610–619
- Vazire S (2010) Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *J Personal Soc Psychol* 98:281–300
- American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, DSM-IV, 4th edn. American Psychiatric Association, Washington, DC
- Achenbach T, Rescorla L (2001) Manual for the ASEBA school-age forms & profiles: an integrated system of multi-informant assessment. University of Vermont, Burlington
- Rey JM, Schrader E, Morris-Yates A (1992) Parent-child agreement on children's behaviours reported by the child behaviour checklist (CBCL). *J Adolesc* 15:219–230

23. Salbach-Andrae H, Lenz K, Lehmkuhl U (2009) Patterns of agreement among parent, teacher and youth ratings in a referred sample. *Eur Psychiatry* 24:345–351
24. Van Widenfelt BM, Goedhart AW, Treffers PD, Goodman R (2003) Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *Eur Child Adolesc Psychiatry* 12:281–289
25. Silverman WK, Albano AM (1996) The anxiety disorders interview schedule for children for DSM-IV: (child and parent versions). Psychological Corporation, San Antonio
26. Lord C, Rutter M, DiLavore PC, Risi S (1999) ADOS. Autism diagnostic observation schedule. manual. Western Psychological Services, Los Angeles
27. Lord C, Rutter M, DiLavore PC, Risi S, Gotham K, Bishop SL (2012). ADOS. Autism diagnostic observation schedule, 2nd edn (ADOS-2). Manual (Part I): modules 1–4. Western Psychological Services, Torrance
28. Schachar R, Wachsmuth R (1989) Parent interview for child symptoms—revised DSM-III-R. Department of Psychiatry, Hospital for Sick Children, Toronto
29. Tannock R, Hum M, Humphries T, Schachar R (2000) Interviewing teachers about children’s classroom behavior and academic performance. In: Proceedings of the 47th annual meeting of the American Academy of Child and Adolescent Psychiatry, New York
30. Feldt LS, Woodruff DJ, Salih FA (1987) Statistical inference for coefficient alpha. *Appl Psychol Meas* 11:93–103
31. Harrell FE Jr (2015) Regression modeling strategies. Springerx, Switzerland
32. Efron B (1986) How biased is the apparent error rate of a prediction rule? *J Am Stat Assoc* 81:461–470
33. Swets JA (1988) Measuring the accuracy of diagnostic systems. *Science* 240:1285–1293
34. Raiker JS, Freeman AJ, Perez-Algorta G, Frazier TW, Findling RL, Youngstrom EA (2017) Accuracy of Achenbach scales in the screening of attention-deficit/hyperactivity disorder in a community mental health clinic. *J Am Acad Child Adolesc Psychiatry* 56:401–409
35. DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44:837–845
36. R Core Team (2016) R: a language and environment for statistical computing, Vienna, Austria
37. Harrell Jr FE (2017) RMS: regression modeling strategies, Vienna, Austria
38. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-, Müller M (2011) pROC: an open-source package for R and S + to analyze and compare ROC curves. *BMC Bioinform* 12:77
39. Fischer JE, Bachmann LM, Jaeschke R (2003) A readers’ guide to the interpretation of diagnostic test properties: clinical example of sepsis. *Intensive Care Med* 29:1043–1051
40. Martel MM, Markon K, Smith GT (2017) Research review: multi-informant integration in child and adolescent psychopathology diagnosis. *J Child Psychol Psychiatry* 58:116–128
41. Goodman R, Renfrew D, Mullick M (2000) Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *Eur Child Adolesc Psychiatry* 9:129–134
42. Christensen DL, Baio J, Van Naarden Braun K, Bilder D, Charles J, Constatino JN, Daniels J, Durkin MS, Fitzgerald RT, Kurzius-Spencer M, Lee L, Pettygrove S, Robinson C, Schulz E, Wells C, Wingate MS, Zahorodny W, Yeargin-Allsopp M (2016) Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2012. *MMWR Surveill Summ* 65:1–23
43. Greene RW, Biederman J, Zerwas S, Monuteaux MC, Goring JC, Faraone SV (2002) Psychiatric comorbidity, family dysfunction, and social impairment in referred youth with oppositional defiant disorder. *Am J Psychiatry* 159:1214–1224
44. Kattan MW, Marasco J (2010) What is a real nomogram? *Semin Oncol* 37:23–26
45. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, Lijmer JG, Moher D, Rennie D, De Vet HC (2015) STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Radiology* 277:826–832
46. Bossuyt PM, Reitsma JB, Bruns D, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, De Vet HC (2003) Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Clin Chem Lab Med* 41:68–73
47. Jensen-Doss A, Youngstrom EA, Youngstrom JK, Feeny NC, Findling RL (2014) Predictors and moderators of agreement between clinical and research diagnoses for children and adolescents. *J Consult Clin Psychol* 82:1151
48. Rettew DC, Lynch AD, Achenbach TM, Dumenci L, Ivanova MY (2009) Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. *Int J Methods Psychiatric Res* 18:169–184