Article

Validity Aspects of the Strengths and Difficulties Questionnaire (SDQ) Adolescent Self-Report and Parent-Report Versions Among Dutch Adolescents

Jorien Vugteveen[1] [iD], Annelies de Bildt[2,3] [iD], Meinou Theunissen[4], Sijmen A. Reijneveld[4,5], and Marieke Timmerman[1]

Abstract

In this study, validity aspects of the Strengths and Difficulties Questionnaire (SDQ) self-report and parent-report versions were assessed among Dutch adolescents aged 12 to 17 years (community sample: $n = 962$, clinical sample: $n = 4,053$). The findings mostly support the continued use of both SDQ versions in screening for psychosocial problems as (a) exploratory structural equation analyses partially supported the grouping of items into five scales; (b) investigation of associations between scales of the SDQ and the Child Behavior Checklist, Youth Self-Report, and Intelligence Development Scales-2 provided evidence for the SDQ versions' convergent and divergent validity; and (c) receiver operating characteristics curves yielded evidence for both SDQ versions' criterion validity by showing that these questionnaires can be used to screen for psychosocial problems, except for the adolescent-reported version for males. Regardless of the adolescent's gender, the receiver operating characteristics curves showed both SDQ versions to be useful for screening for three specific types of problems: anxiety/mood disorder, conduct/oppositional deviant disorder, and attention-deficit/hyperactivity disorder. Additionally, parent-rated SDQ scores were found to be useful for screening for autism spectrum disorder.

Keywords

Psychosocial problems frequently occur in adolescents, with the prevalence estimated at 15% to 25% (e.g., Fergusson, Horwood, & Lynskey, 1993; Ormel et al., 2015). To screen for these problems in community settings, for example, during large scale general health check-ups, the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997, 1999) is a widely used instrument. The SDQ is particularly suitable for this purpose as it (a) is relatively short; (b) focuses on strengths (prosocial behavior) as well as multiple types of difficulties (emotional problems, conduct problems, hyperactivity/inattention, peer problems); and (c) is available in multiple informant versions (self-report, parent, teacher). Of the informant versions, the teacher version is least likely to be relevant for use among adolescents, because adolescents spend only a limited amount of time with each of their teachers. To be of use for screening purposes in an adolescent community population, the SDQ should be of good validity for these populations. As relatively few studies examined the SDQ's validity among adolescents, the purpose of this study was to examine a broad range of validity aspects of the SDQ adolescent self-report and parent versions among Dutch

adolescents. That is, we considered evidence for their presumed internal structure, and their convergent, discriminant, and criterion validity.

Internal Structure

The SDQ was designed to measure strengths as well as four types of difficulties, resulting in a presumed five-factor structure. For the SDQ *adolescent* version, this five-factor

[1]Heymans Institute for Psychological Research, University of Groningen, the Netherlands
[2]Department of Psychiatry, University Medical Center Groningen (UMCG), University of Groningen, the Netherlands
[3]Accare Child and Adolescent Psychiatry Groningen, the Netherlands
[4]Netherlands Organisation of Applied Scientific Research (TNO), Leiden, the Netherlands
[5]Department of Health Sciences, University Medical Center Groningen (UMCG), University of Groningen, the Netherlands

Corresponding Author:
Jorien Vugteveen, Heymans Institute for Psychological Research, University of Groningen, Groningen, Netherlands.
Email: j.vugteveen@rug.nl

structure showed to be tenable in some studies among adolescents (Goodman, 2001; Lundh, Wångby-Lundh, & Bjärehed, 2008; Richter, Sagatun, Heyerdahl, Oppedal, & Røysamb, 2011; Ruchkin, Koposov, & Schwab-Stone, 2007; Van Roy, Veenstra, & Clench-Aas, 2008), but not in others (Bøe, Hysing, Skogen, & Breivik, 2016; Giannakopoulos et al., 2009; Koskelainen, Sourander, & Vauras, 2001; Ortuño-Sierra, Fonseca-Pedrero, Paino, Sastre i Riba, & Muñiz, 2015; Rønning, Handegaard, Sourander, & Mørch, 2004; Van de Looij-Jansen, Goedhart, de Wilde, & Treffers, 2011). It is important to note that none of the studies mentioned can be compared directly with the others, because they strongly differ concerning, for instance, sample age range and country of origin. Another study found a six-factor solution to fit, rather than the five-factor solution (Van Roy et al., 2008). This six-factor structure includes the presumed five factors and an additional *positive construal method* factor. The additional factor consists of the positively worded items, five in total, from the four difficulties scales, implying that this factor expresses the positive wording effects for items measuring difficulties. Note that the positive construal method factor in this six-factor model differs from the positive construal method factor in the modified five-factor model assessed by Van de Looij-Jansen et al. (2011). In their model, the prosocial behavior factor was modified by adding cross-loadings onto the five positively worded items measuring difficulties. By doing so, they ignored that, besides their positive wording, the items measuring prosocial behavior are presumed to have in common that they measure strengths. The resulting factor thus represents a combination of a wording effect and prosocial behavior, implying it is not just a wording factor. For the SDQ *parent* version, the few studies that were conducted found support for the presumed five-factor structure (He, Burstein, Schmitz, & Merikangas, 2013; Van Roy et al., 2008).

Convergent and Discriminant Validity

In previous studies, the SDQ's convergent validity has been investigated using the empirically based syndrome scales of the parent-reported Child Behavior Checklist (CBCL; Achenbach, 1991a) and its self-report version, the Youth Self-Report (YSR; Achenbach, 1991b), as gold standards. Like the SDQ, the CBCL and YSR belong to the domain of instruments measuring behavior, and their validity is well documented (e.g., Achenbach, 1991a, 1991b; Chen, Faraone, Biederman, & Tsuang, 1994; Nakamura, Ebesutani, Bernstein, & Chorpita, 2009; Van Lang, Ferdinand, Oldehinkel, Ormel, & Verhulst, 2005).

Concerning the SDQ's convergent validity, only a few studies were conducted among populations consisting of only adolescents. For the *SDQ adolescent version*, moderate to strong correlations between conceptually similar SDQ and YSR scales were found (Van Widenfelt, Goedhart,

Treffers, & Goodman, 2003; Vogels, Siebelink, Theunissen, Wolff, & Reijneveld, 2011). For the *SDQ parent version*, the only study among adolescents we found, showed moderate correlations between conceptually similar scales of the two instruments (Vogels et al., 2011). Note that the aforementioned studies differed in which of the 11 CBCL/YSR empirically based syndrome scale(s) they regarded as conceptually similar to each SDQ scale. One of the studies compared all SDQ scales with only the three broadband scales (i.e., externalizing problems: delinquent and aggressive behavior; internalizing problems: anxious/depressed, somatic complaints, withdrawn; total problems: sum of all problem items; Vogels et al., 2011), thereby generating only generic results. The other studie additionally considered the eight specific scales (e.g., aggressive behavior, anxious/depressed) by linking each SDQ scale to one or more (Van Widenfelt et al., 2003) syndrome scales.

Of the studies aforementioned, only Van Widenfelt et al. (2003) considered an aspect of discriminant validity. They did so by reporting correlations between conceptually unrelated SDQ and CBCL/YSR syndrome scales. However, whether the convergent correlations (i.e., correlations between scores on related scales) were stronger than the discriminant correlations (i.e., correlations between scores on unrelated scales) was not tested. Note that all scales within a domain can be expected to be associated to some extent, because of the shared domain; conceptually related scales can be expected to be strongly associated, whereas associations among conceptually unrelated scales are expected to be weak.

We were not able to find studies that address the SDQ's discriminant validity by looking at associations between SDQ scales and scales from instruments belonging to unrelated domains, such as the domain of intelligence. Comparing scales across domains is useful because valid measurements of these different domains are expected to show weak or negligible associations.

Criterion Validity

In the few studies we found among adolescent clinical and community samples, the SDQ's ability to distinguish between these two types of samples was found to be good for both the *SDQ adolescent version* (Goodman, Meltzer, & Bailey, 1998; Vogels et al., 2011) and the *SDQ parent version* (Vogels et al., 2011).

Addressing the issues aforementioned, the aim of our study is to examine the internal structure and the convergent, discriminant, and criterion validity of the SDQ adolescent self-report and parent versions among 12- to 17-year-old Dutch adolescents, when used for screening purposes.

First, we will assess both SDQ versions' factor structures among the community sample of adolescents, because we

aim to evaluate the SDQ as it is used in screening. This screening setting resembles the context in which the data were collected, that is, in a community setting. Note that in a previous study using the same data, the SDQ's measurement invariance across clinical and community populations was supported (Vugteveen, de Bildt, Serra, de Wolff, & Timmerman, 2020), which assures us that we do not unintentionally ignore a potential setting effect by looking at only the community data. Here, first we will assess the presumed five-factor structure of both SDQ versions using confirmatory factor analysis (CFA), because this structure most closely resembles how SDQ scale scores are calculated in practice. In case the five-factor structure shows insufficient fit, the fit of a six-factor structure containing the presumed five-factor and a *positive construal methods factor* will be evaluated. These two structures express that the items are perfect indicators of a single (or two) construct(s). As this rarely holds for psychological scales (Asparouhov & Muthén, 2009), we supplement the CFA results with a more exploratory approach, exploratory structural equation modeling (ESEM; Asparouhov & Muthén, 2009). As far as we know, ESEM has only been used on self-reported SDQ scores in one adolescent sample (Garrido et al., 2020), which yielded some support for the presumed five-factor structure, but also indicated items to contribute to scales other than their presumed scale. As further ESEM-based evidence is lacking, we are unsure of whether the presumed five-factor structure will be supported in our study.

Second, the SDQ versions' convergent and discriminant validity will be tested by investigating associations between the SDQ scales and conceptually similar CBCL/YSR scales (same domain), conceptually different CBCL/YSR scales (same domain), and conceptually different Intelligence and Development Scales (IDS-2, different domain; Grob, Meyer, & Hagmann-von Arx, 2018). Considering the results from previous research, we expect to find evidence supporting the SDQ versions' convergent and divergent validity.

Third, we will assess the SDQ scales' ability to distinguish clinical groups from a community group, therewith focusing on the use of the SDQ in a screening context. This clearly differs from an earlier analysis of the clinical data used in this study, where the data were used to investigate how well SDQ scales scores of adolescents referred to mental health care can be used to predict specific types of disorders when used in a clinical context (Vugteveen, de Bildt, Hartman, & Timmerman, 2018). Here, we expect to find support for the use of both SDQ versions' total difficulties scale for distinguishing between the two general groups (community, clinical). Furthermore, as no substantial research is available on how well the each of the five SDQ difficulties and strengths scales can be used to distinguish clinical groups with specific types of disorders from the community group, we have no hypotheses on this matter and we regard our investigation to be exploratory.

Method

Participants

Community Sample. The community sample data of 12- to 17-year-old Dutch adolescents were collected in two waves. The first wave of data was collected in 2009/2010 at secondary schools, if possible as part of a routine well-child care check which is provided to all Dutch adolescents during their second year in secondary education (13- or 14-year-olds). For the 519 adolescents from this wave, adolescent self-reported data ($n = 217$), parent-reported data ($n = 28$), or both ($n = 274$) were available. Also available were YSR data ($n = 211$), CBCL data ($n = 26$), or both ($n = 276$). The second wave of data was gathered in 2016 and 2017 as part of a norming study of an intelligence test, resulting in adolescent self-reported SDQ data ($n = 220$), parent-reported SDQ data ($n = 17$), or both ($n = 206$) from 443 adolescents. Furthermore, YSR data ($n = 181$), CBCL data ($n = 1$), or both ($n = 192$) were available for these adolescents. Additionally, IDS-2 data ($n = 220$) were gathered. Combining data from the two waves resulted in a community sample consisting of 962 adolescents, for whom adolescent-reported SDQ data ($n = 437$), parent-reported SDQ data ($n = 45$), or both ($n = 480$) were available. Also available for the adolescents in this sample were YSR data ($n = 392$), CBCL data ($n = 27$), or both ($n = 468$), and IDS-2 data ($n = 220$). Table S1 (Supplementary Material available online) provides an overview of the available questionnaires within the community sample. The mean age in this sample was 14.1 years ($SD = 1.4$) among males (49.6%) and 14.2 years ($SD = 1.3$) among females (50.4%).

Clinical Sample. The 12- to 17-year-old adolescents in the clinical sample were referred for the first time to one of 29 clinics of an institution for child and adolescent psychiatry in the North of the Netherlands, between January 1, 2013 and December 31, 2015. Their data were collected online during the intake assessment as part of routine outcome monitoring. Of the 4,053 adolescents in the clinical sample, 2,812 had received a *Diagnostic and Statistical Manual of Mental Disorders–Fourth edition* (*DSM-IV*) diagnosis in any of the four categories that content-wise correspond to the SDQ scales. Table S2 (Supplementary Material available online) provides an overview of these diagnoses and an indication of comorbidity of disorders within the sample. The diagnoses were established by trained professionals in a multidisciplinary team, generally consisting of at least a child and adolescent psychiatrist and a child psychologist, and, depending on the context, supplementary professionals such as a specialized nurse. Within this sample, adolescent-reported SDQ data ($n = 354$), parent-reported SDQ data ($n = 206$), or both ($n = 3,493$) were available. The mean age was 14.2 years ($SD = 1.6$) among males (47.6%), and 14.6 years ($SD = 1.5$) among females (52.4%).

Table 1. Demographic and Geographic Characteristics of the Adolescents in the Clinical ($n = 4{,}053$) and Community ($n = 962$) Samples.

Characteristics	Clinical sample, n (%)[a]	Community sample, n (%)[a]	Dutch population, %
Gender			
Male	1,902 (47.6)[b]	474 (49.6)[c]	49.5
Female	2,093 (52.4)	482 (50.4)	50.5
Age, years			
12	581 (14.3)	56 (5.9)[d]	16.5
13	741 (18.3)	315 (33.1)	16.3
14	767 (18.9)	281 (29.5)	16.4
15	799 (19.7)	117 (12.3)	16.9
16	678 (16.7)	107 (11.2)	16.9
17	487 (12.0)	77 (8.1)	17.1
Mother's country of birth			
The Netherlands	[e]	754 (83.2)[f]	78.6
Other	[e]	149 (16.5)	21.4
Mother's educational level			
Low	[e]	187 (24.9)[g]	23.6
Medium	[e]	281 (37.5)	41.7
High	[e]	282 (37.6)	34.7
Geographical region of the Netherlands			
North	2,565 (63.4)[h]	51 (6.9)[i]	10.2
East	1,452 (35.9)	164 (22.2)	21.1
South	4 (0.1)	155 (20.9)	21.4
West	24 (0.6)	367 (49.9)	47.3

[a]Percentages computed of valid cases only. [b]Missing: $n = 58$. [c]Missing: $n = 6$. [d]Missing: $n = 9$. [e]Information not available. [f]Missing: $n = 100$. [g]Missing: $n = 212$. [h]Missing: $n = 10$. [i]Missing: $n = 222$.

Additional demographic and geographic characteristics of both samples are presented in Table 1. For comparison, summary statistics of the Dutch population are presented in the last column of the table (Statistics Netherlands, 2015).

Measures

The Strengths and Difficulties Questionnaire. The 25-item Dutch versions of the SDQ adolescent- and parent-reported versions (Van Widenfelt et al., 2003) both consist of four five-item scales focusing on difficulties relating to emotional functioning, conduct, hyperactivity, and interaction with peers. These four scales together form the total difficulties scale. Additionally, the SDQ contains a five-item scale focusing on strengths in the form of prosocial behavior (Goodman, 1997). The items are rated on a 3-point rating scale (0 = *not true*, 1 = *somewhat true*, and 2 = *certainly true*). Five positively worded items belonging to different SDQ difficulties scales are reverse coded. High scores on the four difficulties scales, represent a high degree of difficulties; a high score on the prosocial scale represents a high degree of prosocial behavior.

The Child Behavior Checklist and Youth Self-Report. The Dutch versions of the CBCL and YSR contain 113 and 112 items, respectively (Verhulst, Van der Ende, & Koot, 1996, 1997).

The items are rated on a 3-point rating scale (0 = *not true*, 1 = *somewhat or sometimes true*, and 2 = *very true or often true*; Achenbach, 1991a, 1991b). For both instruments, all but 17 (CBCL) or 10 items (YSR) can be divided into 8 empirically based syndrome scales with item numbers varying from 8 to 17 (YSR) or 18 (CBCL): (a) aggressive behavior, (b) anxious/depressed, (c) attention problems, (d) delinquent behavior, (e) somatic complaints, (f) social problems, (g) thought problems, (h) withdrawn. Five of these scales can be summarized in two broader scales: (a) the delinquent behavior and aggressive behavior scales form the externalizing behavior scale and (b) the withdrawn, somatic complaints and anxious/depressed scales are combined in the internalizing behavior scale. Together all items, including the items not belonging to the empirically based syndrome scales, form the total behavior problems scale. A second way to summarize 55 of the CBCL and 53 of the YSR items is by dividing them into six *DSM*-oriented scales: (a) affective problems, (b) anxiety problems, (c) attention/deficit/hyperactivity problems, (d) conduct problems, (e) oppositional defiant problems, and (f) somatic problems (Achenbach, 2014).

The Intelligence and Development Scales. The Dutch version of the IDS-2 (Grob, Hagmann-von Arx, Ruiter, Timmerman, & Visser, 2018) contains measures of general

intelligence and of five developmental domains. General intelligence is measured with 14 subtests aimed at visual processing, long-term memory, processing speed, short-term memory (auditory), short-term memory (spatial-visual), abstract thinking, and verbal thinking. The five developmental domains are measured with between two and four subtests per domain, including dividing attention (executive functioning), visual motor skills (psychomotor skills), recognizing emotions (socioemotional competences), logical–mathematical thinking (school skills), and conscientiousness (motivation). All scales are normed, with the general intelligence scale expressed as IQ scores (i.e., $\mu = 100$, $\sigma = 15$) and the five developmental domains as standardized scores (i.e., $\mu = 10$, $\sigma = 3$).

Statistical Analysis

Missing Data. Our data set contained missing data at two levels: questionnaire level and item level. First, for some participants entire SDQ, CBCL, YSR, or IDS-2 questionnaires were unavailable resulting in missing data at questionnaire level. The sample description of both samples contains information about the available questionnaires. Second, the community sample data set contained some missing data at item level for the SDQ adolescent version ($M = 0.33\%$, $SD = 0.32$, min $= 0.0\%$, max $= 1.2\%$) and the SDQ parent version ($M = 0.38\%$, $SD = 0.28$, min $= 0.0\%$, max $= 0.8\%$). This sample data set further contained some missing data at item level for the YSR within the group of adolescents that also filled in the SDQ ($M = 0.69\%$, $SD = 0.50$, min $= 0.1\%$, max $= 4.4\%$); and for the CBCL within the group of parents that filled in the SDQ ($M = 0.85\%$, $SD = 0.$ 53, min $= 0.2\%$, max $= 4.2\%$). The missing data at questionnaire level was not imputed; analyses were performed based on available cases. Taking into account the small number of missing values at item level and the type of analyses we were planning to perform, these missing data were imputed in two ways. First, for the calculation of SDQ, YSR, and CBCL scale scores, mean imputation of item scores was used, in compliance with the instruments' manuals. For the CBCL and the YSR, five parents and four adolescents had too many scores missing to calculate a score for the *DSM*-oriented somatic problems scale; these item scores were not imputed, resulting in missing scale scores. All other missing scores were imputed. The resulting scale scores were used for analyses at scale level based on available cases: calculating mean scale scores and correlations between scale scores. Second, for analyses at item level, two-way imputation with normally distributed errors was used to impute the missing data (e.g., Van Ginkel, Ark, & Sijtsma, 2007); this approach, unlike mean imputation, leads to unbiased item covariance estimates, which is preferred for item level analyses. The two-way imputed data were used for confirmatory factor analyses on the SDQ data and estimating the reliability of the SDQ, CBCL and YSR scales.

Among the adolescents in the community sample that had IDS-2 data available, some IDS-2 data were missing at domain level ($M = 4.32\%$, $SD = 3.48$, min $= 0.0\%$, max $= 10.0\%$). Underlying are missing data at subtest level. We deemed it unwise to impute entire subtests and decided to perform the analyses regarding the IDS-2 data based on available cases.

Factor Structure. The factor structures of the SDQ versions (adolescent, parent) were evaluated using the community sample data. Per SDQ version, the presumed five-factor structure was modeled using CFA for ordinal data (B. Muthén, 1984). The CFA models were estimated using weighted least squares mean and variance adjusted (WLSMV) estimation. Goodness of fit was assessed by considering the comparative fit index (CFI; Bentler, 1990) and the root mean square error of approximation value (RMSEA; Steiger, 1980). We consider CFI values $\geq.90$ combined with RMSEA values $\leq.08$ to be acceptable, while preferring CFI values $\geq.95$ combined with RMSEA values $\leq.06$ (Hu & Bentler, 1999; Marsh, Hau, & Wen, 2004). For comparability with other studies, Tucker–Lewis index values were also presented (Tucker & Lewis, 1973). In case the RMSEA and CFI values indicated insufficient fit of the five-factor model, the six-factor alternative was evaluated. This factor structure consists of the presumed five factors and an additional positive construal method factor containing five positively worded items from the four difficulties scales. The positively worded items of the prosocial behavior scale were not included in this additional factor as these items differ from the five positively worded items measuring difficulties. They differ from each other in the sense that the prosocial items indicate a strength and jointly make up a single scale that does not contain any negatively worded items, whereas the positively worded items difficulties items from the positive construal method factor are part of scales that contain both positively and negatively worded items.

One of the main characteristics of CFA is that it only allows items to load on the factor(s) they are presumed to contribute to, and it fixes other cross-loadings at 0. In our model this implies that each item has a freely estimated loading on a single factor only. Although this closely resembles how SDQ scale scores are calculated in practice, it may distort model fit (Marsh, Morin, Parker, & Kaur, 2014) and inflate associations between factors, which in turn affects the estimated factor loadings and factor reliabilities (e.g., Asparouhov, Muthén, & Morin, 2015). To overcome these limitations, we supplemented our analyses with ESEM using WLSMV estimation and target rotation (Asparouhov & Muthén, 2009; Marsh et al., 2014). The latter aims to minimize cross-loadings without forcing them to be 0. As

with CFA, we used ESEM to test the fit of the presumed five-factor structure. In case that model did not fit, we evaluated the fit of the six-factor structure. For all factor analyses, loadings ≥.30 are regarded as salient loadings.

For CFA and ESEM models that showed sufficient fit, local fit was assessed (Supplementary Material available online) using the standardized expected parameter change statistic (SEPC; Saris, Satorra, & van der Veld, 2009). SEPC values >.20 warranted allowing item residuals to correlate by freeing them one at the time, starting with the parameter associated with the largest SEPC, until acceptable local fit was found.

Scale Reliabilities. Per SDQ scale, the reliability of the observed scores was computed using the nonlinear structural equation modeling reliability coefficient (ρ_{NL}; Yang & Green, 2015), based on a one-factor model including correlated item residuals as far as necessary to achieve acceptable local fit, as indicated by SEPC values. The reliability coefficient takes into account both the SDQ items' ordinal nature and allows for unequal item loadings per factor (non-tau-equivalence). SDQ scales were considered sufficiently reliable when $\rho_{NL} \geq .70$, while ≥.80 was preferred (Evers, Sijtsma, Lucassen, & Meijer, 2010). For the purpose of comparability with other studies, Cronbach's alpha coefficients were calculated for all SDQ, CBCL, and YSR scales. For the IDS-2, we lacked the item scores necessary to compute Cronbach's alpha.

The analyses mentioned so far are analyses performed at item-level. For the remaining analyses, scale-level data were used.

Descriptive Statistics. To characterize differences across informants and settings, mean scale scores were calculated per SDQ, CBCL, and YSR scale. Note that SDQ scores were available for both settings (community, clinical), and all other instruments for the community setting only. In contrast to SDQ, CBCL, and YSR scores, IDS-2 scores were normed, allowing us to compare community scores with population means. For this purpose, *z* tests were used. To assess potential setting differences in SDQ scale scores per SDQ version, a multivariate analysis of variance (MANOVA) with the SDQ scale scores as dependent variables and the setting as independent variable was conducted, followed by *t* tests for post hoc univariate comparisons per SDQ version and scale to compare scale scores across settings. Given the nature of the populations, it is to be expected that the prevalence of psychiatric disorders related to psychosocial functioning was higher in the clinical sample than in the community sample. Therefore, we expect to find higher mean scale scores for the SDQ difficulties scales and a lower mean scale score for the SDQ strength scale.

Convergent and Discriminant Validity. To express the strength of associations of rank scores on SDQ (adolescent, parent) and YSR (adolescent)/CBCL (parent) scale pairs, we computed Spearman Rho correlations. These correlations were computed for conceptually related SDQ and YSR/CBCL scale pairs, denoted as convergent correlations, and with conceptually different SDQ and CBCL/YSR or IDS-2 scale pairs, denoted as discriminant correlations. Per SDQ scale, Steiger's (1980) test was used to compare convergent with discriminant correlations within the set of (a) eight empirically based syndrome scales, (b) eight empirically based syndrome scales and the three broader empirically based syndrome scales, and (c) six *DSM*-oriented scales.

Criterion Validity. In order to determine how well both SDQ versions were able to distinguish between the community and clinical populations, we used receiver operating characteristic (ROC) curves. First, we investigated how well the SDQ total difficulties scale of both SDQ versions was able to distinguish between the two populations. Next, we examined each SDQ difficulties and strengths scale's ability to differentiate between the community population and a clinical subpopulation that had received a diagnosis content-wise corresponding to the particular SDQ scale (anxiety/mood disorder for the SDQ emotional scale, conduct/oppositional deviant disorder [CD/ODD] for the SDQ conduct scale, attention-deficit/hyperactivity disorder [ADHD] for the SDQ hyperactivity scale, and autism spectrum disorder [ASD] for the SDQ social problems and prosocial behavior scales). Additionally, we provided an investigation into potential gender differences (Supplementary Material available online). Area under the curve (AUC) values were reported as an index of discriminative ability. We considered AUC values ≥.80 as indicating sufficient ability to distinguish between samples. For comparing AUC values of different SDQ scales, DeLong's test for paired ROC curves was used (DeLong, DeLong, & Clarke-Pearson, 1988).

For all statistical tests, a significance level of α = .01 was used. The confirmatory factor and ESEM analyses were performed in M*plus* version 8.0 (L. K. Muthén & Muthén, 1998-2017). All other analyses were performed in R, version 3.4.1 (R Core Team, 2016). Data imputation was performed using the Mokken package (Van der Ark, 2007), the ROC analyses were performed using the pROC package (Robin et al., 2011), and the ρ_{NL} coefficients were computed using the semTools package (Jorgensen, Pornprasertmanit, Schoemann, & Rosseel, 2018).

Results

Factor Structure of SDQ Adolescent and Parent Versions

Table 2 presents the goodness-of-fit statistics of the CFA and ESEM models evaluated using community sample data. For *the adolescent version*, the CFA models showed insufficient fit for the five-factor model and acceptable fit for the

Table 2. Goodness-of-Fit Statistics of the CFA and ESEM Models for the SDQ Adolescent and Parent Versions in the Community Sample.

Model	χ^2	df	p	RMSEA	RMSEA, 90% CI	CFI	TLI
SDQ adolescent version							
CFA-5F	772.988	265	<.001	.046	[.042, .049]	.896	.883
CFA-6F	525.249	255	<.001	.034	[.030, .038]	.945	.935
ESEM-5F	304.576	185	<.001	.027	[.021, .032]	.976	.960
SDQ parent version							
CFA-5F	576.368	265	<.001	.047	[.042, .053]	.926	.916
ESEM-5F	274.950	185	<.001	.030	[.023, .038]	.979	.965

Note. df = degrees of freedom; RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker–Lewis index; SDQ = Strengths and Difficulties Questionnaire; CFA = confirmatory factor analysis; ESEM = exploratory structural equation modeling. For the SDQ adolescent version, *n* = 917 and for the SDQ parent version, *n* = 525.

six-factor model, suggesting the potential presence of a wording effect. As both CFA models still may have misrepresented the SDQ's factor structure, the five-factor ESEM model was evaluated. This model showed excellent fit. Table 3 presents factor loadings and factor correlations for both CFA models and the ESEM model. Note that two items in the ESEM model (Items 7 "obedient" and 11 "friend", both positively worded items measuring difficulties) showed negligible loadings on their intended factor (loadings ≤ .30) and one item (Item 1 "considerate", prosocial factor) loaded on its intended factor as well as on the conduct difficulties factor.

Information about the local fit of the six-factor CFA model and the five-factor ESEM model is provided in Tables S3 and S4 (Supplementary Material available online). Per model, three error correlations were added to the model, indicating that three item pairs formed subfactors within the factor they belong to. One additional item (Item 5 "temper", conduct factor) now showed substantial loadings on its intended factor as well as on the emotional difficulties factor.

For *the parent version*, the five-factor CFA model fitted acceptably; the five-factor ESEM model fitted better. Table 4 presents factor loadings and factor correlations for both CFA models and the ESEM model. The ESEM model showed one item (Item 5 "temper", conduct factor) loading negligibly on its intended factor (loading ≤ .30). This item and five other items (Items 10 "fidgety", 14 "generally liked", 17 "kind", 19 "bullied", and 24 "fears") showed salient but weak loadings (loadings ranging from .30 to .37) on a factor they were not intended to load on.

For this SDQ version, information about the local fit of the five-factor CFA and ESEM models is provided in Tables S3 and S5 (Supplementary Material available online). Four error correlations were added to the CFA model, and two were added to the ESEM model, indicating the presence of subfactors. One additional item (Item 12 "temper", conduct factor) now showed a negligible loading on its intended factor.

Scale Reliability

For the SDQ *adolescent* version, ρ_{NL} estimates of .73, .55, .72, .56, and .63 were found for the emotional difficulties, conduct difficulties, hyperactivity/inattention problems, social problems, and prosocial behavior scales, respectively. Regarding the SDQ *parent* version, ρ_{NL} estimates for these scales were .71, .57, .72, .68, and .75. The estimates suggested questionable reliability for four out of five adolescent-reported SDQ scales and two out of five parent-reported SDQ scales. Cronbach's alpha coefficients per scale of both SDQ versions and the CBCL/YSR are presented in Tables 5 and 6, respectively.

Scale Scores

Community setting mean scale scores of both SDQ versions and the CBCL/YSR are presented in Tables 5 and 6, respectively. Note that it is impossible to gain insight into relative problem levels in our sample by comparing mean scale scores within an instrument to each other, because some types of behavior are generally less prevalent than the others. Table 7 provides community setting mean scale scores for the IDS-2. The IDS-2 scales were normed, allowing us compare our sample means with population means. Table 7 presents the outcomes of the *z*-tests that were used. The community sample scored significantly lower than the population on the general intelligence scale, but not on the five developmental domains.

Table 5 additionally presents mean scale scores for both SDQ versions in the clinical setting. The MANOVA and post hoc *t* tests performed to assess potential setting differences in SDQ scale scores per SDQ version, showed significant setting effects on all SDQ scales, except the adolescent-reported prosocial behavior scale, $t(4,762) = 8.26$, $p = .16$, with higher scores on the SDQ difficulties scales, and lower scores on the parent-reported SDQ prosocial scale, in the clinical setting than in the community setting, $F(3, 962) = 120.09$, $p < .001$.

Table 3. Standardized Parameter Estimates of the CFA and ESEM Models for the SDQ Adolescent Version.

Item/ factor	CFA five-factor model					CFA six-factor model						ESEM five-factor model				
	ES	CP	HP	SP	PB	ES	CP	HP	SP	PB	PCM	ES	CP	HP	SP	PB
3	**.49**					**.49**						**.54**	.21	.003	−.17	.04
8	**.72**					**.72**						**.67**	.10	0.02	.07	.17
13	**.79**					**.79**						**.75**	.12	−0.01	.05	.09
16	**.64**					**.64**						**.60**	−.18	.06	.12	−.08
24	**.78**					**.78**						**.72**	−.22	.06	.18	−.06
5		**.72**					**.78**					.25	**.49**	.16	.13	.02
7		**.45**					.05				.36	−.04	**.23**	.21	−.17	−.30
12		**.59**					.61					−.08	**.66**	.05	.03	−.06
18		**.64**					.67					−.13	**.55**	.16	.28	.01
22		**.60**					.62					−.09	**.53**	.02	.07	−.01
2			**.77**					.79				−.16	−.004	**.90**	.13	.13
10			**.73**					.75				.002	.01	**.75**	.13	.15
15			**.77**					.79				.15	.03	**.73**	−.13	−.002
21			**.57**					.35			.40	.05	.21	**.38**	−.14	−.19
25			**.64**					.50			.28	.12	−.01	**.55**	−.20	−.25
6				**.56**					.64			.15	−.11	−.03	**.60**	−.14
11				**.51**					.40		.61	.14	.24	−.09	**.15**	−.27
14				**.71**					.58		.30	.04	.18	.04	**.45**	−.25
19				**.68**					.74			.11	.19	.07	**.57**	.08
23				**.49**					.55			.14	.09	−.08	**.45**	−.01
1					**.77**					.77		.15	−.42	.03	−.07	**.52**
4					**.46**					.45		.01	.01	.03	−.22	**.42**
9					**.62**					.62		.06	.18	−.07	−.15	**.72**
17					**.64**					.63		−.02	−.13	−.05	−.07	**.49**
20					**.53**					.54		−.02	.06	−.02	.10	**.68**
Factor correlations																
ES	1.00	0.28	0.34	0.58	−0.02	1.00	0.33	0.37	0.59	−0.02	−.06	1.00	0.10	0.30	0.41	−0.03
CP		1.00	0.63	0.54	−0.62		1.00	0.52	0.50	−0.52	−.42		1.00	0.38	0.24	−0.34
HP			1.00	0.24	−0.31			1.00	0.17	−0.20	.35			1.00	0.11	−0.23
SP				1.00	−0.45				1.00	−0.28	−.09				1.00	−0.12
PB					1.00					1.00	−.71					1.00
PCM											1.00					

Note. SDQ = Strengths and Difficulties Questionnaire; ESEM = exploratory structural equation modeling; CFA = confirmatory factor analysis; ES = emotional symptoms; CP = conduct problems; HP = hyperactivity/attention problems; SP = social problems; PB = prosocial behavior; PCM = positive construal method. Per item, its loading on its intended factor is printed in bold.

Convergent and Discriminant Validity

Table 8 presents Spearman rho correlations between the SDQ scales of the SDQ parent version and the CBCL (parent-reported) scales, and between the SDQ adolescent version and the YSR (adolescent-reported) scales. Convergent correlations (correlations between conceptually similar scales) are printed in bold; the remaining correlations are discriminant correlations (correlations between conceptually different scales). All but five of the resulting correlations were significantly different from 0, with convergent correlations ranging from .39 to .79 and discriminant correlations from .12 to .68. Per SDQ scale and for all but 13 comparisons, the convergent correlations were positive and significantly stronger than the discriminant correlations, in line with our expectations.

Table 9 presents Spearman rho correlations between the scales of both SDQ versions and the IDS-2 scales. Of the resulting correlations, which are all considered discriminant correlations, only 16 were significantly different from 0. These 16 correlations, ranging from −.38 to −.19, indicated the presence of weak negative relationships between SDQ and IDS-2 scores, which is in line with our expectations. All but four of these correlations were found between scales of the SDQ *adolescent* version and IDS-2 scales, suggesting that adolescent self-reported SDQ scale scores were slightly more, but at most weakly, associated with the adolescent's intelligence than parent-reported scores.

Table 4. Standardized Parameter Estimates of the CFA and ESEM Models for the SDQ Parent Version.

Item/factor	CFA five-factor model					ESEM five-factor model				
	ES	CP	HP	SP	PB	ES	CP	HP	SP	PB
3	**.34**					**.45**	.04	.05	−.19	.02
8	**.84**					**.85**	.04	−.02	.14	.13
13	**.79**					**.78**	−.14	.06	.14	.05
16	**.78**					**.56**	.26	−.01	.21	.04
24	**.78**					**.55**	.35	−.10	.26	.02
5		**.62**				.34	**.17**	.22	−.06	−.10
7		**.57**				.06	**.36**	.17	−.16	−.34
12		**.42**				.08	**.39**	.10	.15	.14
18		**.71**				.10	**.53**	.25	−.12	−.18
22		**.49**				.16	**.66**	−.05	−.08	−.02
2			**.78**			−.25	.16	**.80**	.24	.14
10			**.77**			−.18	.17	**.74**	.34	.24
15			**.86**			.12	−.05	**.84**	−.07	.01
21			**.61**			−.01	.17	**.50**	−.13	−.21
25			**.83**			.18	−.18	**.86**	−.20	−.14
6				**.53**		.19	−.09	−.09	**.41**	−.26
11				**.63**		.03	−.04	.08	**.59**	−.20
14				**.75**		.18	−.06	.13	**.40**	−.37
19				**.80**		.33	.04	.25	**.48**	.05
23				**.66**		.17	−.06	.04	**.58**	−.14
1					**.87**	.01	−.23	−.14	−.13	**.65**
4					**.78**	−.08	−.13	.14	−.23	**.67**
9					**.75**	.15	−.05	.02	−.19	**.77**
17					**.62**	−.01	.31	−.03	−.22	**.65**
20					**.61**	.15	−.16	.01	.14	**.77**
Factor correlations										
ES	1.00	0.51	0.39	0.68	−0.21	1.00	0.19	0.35	0.36	−0.19
CP		1.00	0.67	0.46	−0.54		1.00	0.37	0.17	−0.12
HP			1.00	0.38	−0.28			1.00	0.17	−0.19
SP				1.00	−0.57				1.00	−0.22
PB					1.00					1.00

Note. SDQ = Strengths and Difficulties Questionnaire; ESEM = exploratory structural equation modeling; CFA = confirmatory factor analysis; ES = emotional symptoms; CP = conduct problems; HP = hyperactivity/attention problems; SP = social problems; PB = prosocial behavior. Per item, its loading on its intended factor is printed in bold.

Criterion Validity

The AUC values presented in Table 10 indicate sufficient discriminative ability of all SDQ scales, except for the adolescent-reported social problems scale and the adolescent- and parent-reported prosocial behavior scales. The latter were not corroborated as being insufficiently capable of distinguishing between the community sample and the clinical subsample of adolescents with an ASD diagnosis. It is noteworthy that for both SDQ versions, the emotional difficulties, the conduct problems, and hyperactivity/inattention scales were better at distinguishing between types of disorders than the SDQ total difficulties scale was at distinguishing between the total community and clinical samples. The ROC graphs are provided (Figures S1 to S10, Supplementary Material available online). Table S6, Table S7, and Figures S11 to S30 (Supplementary Material available online) provide an investigation of potential gender effects. The main gender difference was found for the SDQ adolescent version's total difficulties scale, which distinguished sufficiently between the community and clinical samples for females (AUC = .84) but not for males (AUC = .76).

Discussion

The aim of this study was to investigate validity aspects of the SDQ adolescent self-report and parent versions among 12- to 17-year-old Dutch adolescents in a community setting. We focused on the SDQ versions' internal structure, and convergent, discriminant, and criterion validity.

Table 5. Per SDQ Version (Adolescent, Parent) and per Setting (Community, Clinical): Mean Scale Scores, Standard Deviations, and Cronbach's Alpha.

| | | SDQ version | | | |
| | | Adolescent[a] | | Parent[b] | |
Setting	SDQ scale	α[c]	M (SD)	α[c]	M (SD)
Community	Total[c]	.66	8.1 (4.8)	.70	6.4 (5.0)
	Emotional	.68	2.1 (2.0)	.69	1.6 (1.9)
	Conduct	.51	1.3 (1.3)	.46	0.8 (1.2)
	Hyper	.74	3.4 (2.3)	.78	2.4 (2.4)
	Social	.54	1.3 (1.5)	.64	1.5 (1.8)
	Prosocial	.61	8.0 (1.7)	.72	8.3 (1.8)
Clinical	Total	.70	14.5 (5.9)	.67	15.9 (6.5)
	Emotional	.77	4.4 (2.8)	.75	5.0 (2.8)
	Conduct	.58	2.5 (1.8)	.73	2.8 (2.4)
	Hyper	.76	5.3 (2.6)	.76	5.2 (2.8)
	Social	.54	2.3 (1.9)	.66	2.9 (2.3)
	Prosocial	.64	7.9 (1.8)	.74	7.4 (2.2)

Note. SDQ = Strengths and Difficulties Questionnaire; α = Cronbach's index of internal consistency (alpha).
[a]Adolescent version clinical setting, $n = 3,847$; community setting, $n = 3,699$. [b]Parent version clinical setting, $n = 917$; community setting, $n = 525$.
[c]Per SDQ version, all mean scale score comparisons across settings, except the comparison for the adolescent-reported prosocial behavior scale, indicated a significant difference with $p < .001$.

Table 6. For the Adolescent Self-Reported YSR and the Parent Reported CBCL: Mean Scale Scores, Standard Deviations, and Cronbach's Alpha (Community Setting).

| | Informant | | | |
| | Adolescent ($n = 850$) | | Parent ($n = 489$) | |
YSR/CBCL scale	α	M (SD)	α	M (SD)
Empirically based syndrome scales				
Aggressive problems	.81	3.7 (3.8)	.85	2.4 (3.4)
Anxious/depressed	.84	3.5 (3.9)	.80	2.1 (2.8)
Attention problems	.76	4.4 (3.1)	.81	3.0 (3.1)
Delinquent	.69	3.1 (2.8)	.69	1.2 (1.9)
Social problems	.69	2.7 (2.6)	.77	1.4 (2.3)
Somatic complaints	.75	2.6 (2.8)	.63	1.5 (1.9)
Thought problems	.72	2.7 (3.0)	.63	1.4 (2.0)
Withdrawn	.73	2.6 (2.5)	.77	1.8 (2.3)
Total	.93	23.4 (15.7)	.93	13.8 (12.5)
Externalizing	.86	6.8 (6.0)	.87	3.6 (4.8)
Internalizing	.89	8.8 (7.8)	.86	5.4 (5.6)
DSM-oriented scales				
Affective problems	.78	3.3 (3.5)	.72	1.6 (2.3)
Anxiety problems	.66	2.0 (2.0)	.66	1.0 (1.5)
Attention problems	.76	4.2 (2.9)	.81	2.3 (2.6)
Conduct problems	.71	2.5 (2.7)	.71	0.9 (1.7)
Oppositional defiant problems	.63	1.6 (1.6)	.76	1.2 (1.6)
Somatic problems[a]	.68	1.6 (2.0)	.54	1.1 (1.4)

Note. YSR = youth self-report; CBCL = child behavior checklist; α = Cronbach's index of internal consistency (alpha); DSM = *Diagnostic and Statistical Manual of Mental Disorders*.
[a]YSR, $n = 846$ (scale score missing for four cases); CBCL, $n = 484$ (scale score missing for five cases).

Table 7. IDS-2 Mean Scale Scores (Community Setting).

IDS-2	*n*	*M (SD)*
General intelligence	216	93.8 (16.9)[a]
Executive functioning	214	9.9 (2.2)[b]
Psychomotor skills	207	10.5 (2.1)[b]
Socioemotional competences	209	10.3 (3.1)[b]
School skills	215	9.5 (2.7)[b]
Motivation	198	10.4 (3.0)[b]

Note. IDS-2 = Intelligence Development Scale–2; CI = confidence interval.
[a]Significantly different from the normed population means (general intelligence: $z = -6.07$, $p < .001$, 99% CI [91.17, 96.43]). [b]Not significantly different from the normed population means (executive functioning: $z = -0.49$, $p = .626$, 99% CI [9.37, 10.43]; psychomotor skills: $z = 2.40$, $p = .017$, 99% CI [9.96, 11.04]; socioemotional competences $z = 1.45$, $p = .148$, 99% CI [9.77, 10.84]; school skills: $z = -2.44$, $p = .015$, 99% CI [8.97, 10.03]; motivation: $z = 1.88$, $p = .061$, 99% CI [9.85, 10.95]).

Internal Structure

Holding ESEM models in higher regard than CFA models, due to the plausibility of items loading on more than one factor, we found some support for the presumed five-factor structure. However, three items of the SDQ adolescent version and six items of the parent version were found to be somewhat questionable indicators of their theoretical construct, with one (parent version) or two (adolescent version) items failing to substantially contribute to the scale they were presumed to contribute to and some items unexpectedly contributing to other scales than their presumed scale. Additionally, the analyses revealed the presence of two to four correlated residuals per SDQ version that were not intended to exist. Scale score reliabilities were sufficient for the self-reported hyperactivity/inattention scale and for the parent reported emotional difficulties, hyperactivity/inattention, and prosocial behavior scales, but not for the other scales of both SDQ versions. These findings are cause for concern, but can possibly partially be attributed to the fact that the SDQ aims to measure five dimensions of psychosocial functioning with only five items per dimension. The SDQ's briefness, widely considered to be one of its perks, may come at a cost. Additionally, it is worth noticing that the samples used in this study are presumably large enough to obtain accurate results with CFA's. ESEM models, on the other hand, are substantially less parsimonious and thus require larger samples (Garrido et al., 2020), which warrants some caution with regard to the results of our ESEM analyses.

For the adolescent version, our factor structure and reliability findings are in line with findings by Garrido et al. (2020), who performed the only other study using ESEM for assessing the SDQ's scale structure. As none of the other investigations into the factor structure of the adolescent and parent versions are based on ESEM, it is difficult to

compare the findings of the current study with other studies. Our reliability findings appear to deviate from previous research, with most previous studies finding higher reliability estimates than we did. However, note that previous studies have used either Cronbach's alphas or ordinal alphas to estimate reliability, which are both suboptimal measures of the reliability of SDQ scores as Cronbach's alpha does not take the SDQ items' ordinal nature into account and ordinal alpha estimates the reliability of the latent continuous variables underlying the observed scores.

Convergent and Discriminant Validity

Using the CBCL and YSR as gold standards, we found evidence for the SDQ adolescent and parent versions' convergent and discriminant validity as, in the great majority of cases, each SDQ scale was more strongly associated with its conceptually similar CBCL/YSR scale(s) than with conceptually different CBCL/YSR scales. These findings are in line with our expectations and with findings from previous studies (Van Widenfelt et al., 2003; Vogels et al., 2011). Note that the comparison with findings from previous studies is slightly hampered by the fact that these studies differed to some extent with regard to the CBCL/YSR scales they identified as conceptually similar to the SDQ scales. Besides, two out of the three studies did not compare SDQ scales with conceptually different CBCL/YSR scales, therewith impeding a comparison of our outcomes regarding discriminant validity with previous studies.

Compared with the aforementioned previous studies, our study adds two unique perspectives to the investigation of the SDQ's convergent and discriminant validity. First, while previous studies only compared the SDQ scales with the CBCL/YSR empirically based syndrome scales, our study additionally compares the SDQ scales with the CBCL/YSR *DSM*-oriented scales. The *DSM*-oriented scales result from a top-down approach of grouping items based on their coverage of *DSM* symptom categories, whereas the empirically based syndrome scales result from a bottom-up approach of applying statistical analyses to group items. As item grouping based on criteria formulated for diagnostic purposes is clinically relevant, we regard the findings regarding the comparison of the SDQ scales with the *DSM*-oriented CBCL/YSR scales as additional evidence for the SDQ scales' convergent and discriminant validity.

The second perspective, which makes our study standout from previous studies, is that we investigated the SDQ's discriminant validity by comparing SDQ scales to scales of an instrument from a different domain: the IDS-2 from the domain of intelligence tests. We deem this a useful comparison as lack of a shared domain can be expected to result in weak to negligible associations between scales of instruments from different domains. In the current study, this endeavor resulted in additional evidence for the SDQ's

Table 8. Spearman Rho Correlations Between SDQ Scores and YSR/CBCL Scale Scores (Community Setting).

YSR/CBCL scales	Scales SDQ adolescent version[a]					Scales SDQ parent version[a]				
	Total	Emotion	Conduct	Hyper	Social	Total	Emotion	Conduct	Hyper	Social
Empirically based syndrome scales										
Aggressive problems	.55	.33	**.45**	.45	.20	.57	.35	**.59**	.44	.24
Anxious/depressed	.53	**.68**	.13	.25	.27	.42	**.56**	.22	.14	.25
Attention problems	.65	.34	.35[b]	**.72**	.15	.68	.33	.40[b]	**.74**	.23
Delinquent	.45	.20	**.43**	.37	.20	.46	.25	**.48**	.35	.22
Social problems	.56	.47	.24	.33	**.39**	.58	.44	.36	.36	**.43**
Somatic complaints	.47	.51	.18	.29	.17	.29	.45[b]	.14	.11[c]	.15
Thought problems	.56	.45	.28	.40	.29[b]	.44	.36	.26	.34	.21
Withdrawn	.53	.55	.16	.22	**.47**	.51	.41	.21	.21	**.54**
Externalizing	.57	.31	**.49**	.47	.22	.59	.36	**.60**	.45	.25
Internalizing	.62	**.71**	.18	.31	.36[d]	.54	**.61**	.25	.21	.42[b]
Total	**.74**	.58	.40[d]	.55	.32	**.73**	.54[b]	.50[d]	.54	.36
DSM-oriented scales										
Affective problems	.60	.56[e]	.26	.38	.34	.51	.45[e]	.31	.30	.33
Anxiety problems	.51	**.62**	.12	.26	.26	.44	**.53**	.22	.22	.23
Attention problems	.58	.24	.35[e]	**.74**	.05[c]	.67	.30	.41[c]	**.79**	.16
Conduct problems	.44	.19	**.42**	.37	.17	.45	.23	**.52**	.36	.18
Oppositional defiant problems	.45	.25	**.43**	.36	.16	.50	.28	**.55**	.39	.19
Somatic problems[f]	.38	.43	.14	.23	.12	.23	.41	.11[c]	.08[c]	.08[c]

Note. SDQ = Strengths and Difficulties Questionnaire; YSR = youth self-report; CBCL = child behavior checklist. Correlations between conceptually similar scales (convergent correlations) are presented in bold. Unlike the other discriminant correlations, this discriminant correlation is not significantly stronger than the lowest of the convergent correlations between the associated SDQ scale and each of the eight empirically based CBCL/YSR scales, all empirically based CBCL/YSR scales, or the *DSM*-oriented CBCL/YSR scales.
[a]SDQ adolescent version: YSR combination, *n* = 840; SDQ parent version: CBCL combination, *n* = 456. [b]Empirically based CBCL/YSR scales.
[c]Correlation not significant at the .01 level; all other correlations are significant at the .01 level. [d]All empirically based CBCL/YSR scales. [e]The *DSM*-oriented CBCL/YSR scales. [f]YSR, *n* = 836 (four cases missing); CBCL, *n* = 451 (five cases missing).

Table 9. Spearman Rho Correlations Between SDQ Scores and IDS-2 Scale Scores (Community Setting).

IDS-2 scales	Scales SDQ adolescent version						Scales SDQ parent version				
	n	Total	Emotional	Conduct	Hyper	Social	*n*	Total	Emotional	Conduct	Hyper	Social
General intelligence	204	−.20*	.01	−.31*	−.01	−.33*	137	−.32*	−.15	−.21	−.19	−.30*
Developmental domains												
Executive functioning	202	−.15	.00	−.23*	.00	−.26*	136	−.21	−.12	−.06	−.10	−.27*
Motivation for school	187	−.28*	−.10	−.18	−.38*	.01	127	−.10	.07	−.14	−.19	−.01
Psychomotor skills	195	−.17	−.11	−.10	−.12	−.09	131	−.18	−.13	−.05	−.16	−.08
School skills	203	−.20*	−.07	−.24*	−.03	−.29*	136	−.24*	−.17	−.12	−.14	−.22
Socioemotional competences	197	−.19*	.06	−.28*	−.14	−.19*	134	−.08	−.10	−.08	−.04	−.20

Note. SDQ = Strengths and Difficulties Questionnaire; IDS = Intelligence Development Scales.
*Correlation significant at the .01 level.

discriminant validity as scores on SDQ and IDS-2 scales appeared to be unrelated or weakly negatively related to each other.

To summarize, our findings suggest that the SDQ measures the intended four types of difficulties and does not unintendedly measure other aspects of behavior or intelligence.

Criterion Validity

For *both* SDQ versions, our findings indicate that the SDQ total difficulties scale can be used to distinguish between community and clinical populations, as is in line with conclusions drawn in previous studies (Goodman et al., 1998; Vogels et al., 2011). In other words, in a screening context,

Table 10. Per SDQ Version and Scale, Its Ability to Distinguish Between Community and Clinical (Sub)Samples.

	SDQ version					
	Adolescent			Parent		
SDQ scale	Comm., *n*	Clin., *n*[a]	AUC (*SE*)	Comm., *n*	Clin., *n*	AUC (*SE*)
Total	917	3,847	.80 (.01)	525	3,699	.87 (.01)
Emotional	917	1,325	.87 (.01)	525	1,215	.92 (.01)
Conduct	917	363	.85 (.01)	525	346	.93 (.01)
Hyper	917	873	.85 (.01)	525	856	.91 (.01)
Social	917	667	.75 (.01)	525	670	.84 (.01)
Prosocial	917	667	.58 (.01)	525	670	.75 (.01)

Note. SDQ = Strengths and Difficulties Questionnaire; Comm. = community sample; Clin. = clinical (sub)sample; AUC = area under the curve; *SE* = standard error.
[a]Per SDQ scale, the clinical subsamples consisted of adolescent with a *DSM-IV* diagnosis content-wise matching the SDQ scale: Anxiety/Mood disorder for the SDQ emotional scale, Conduct/Oppositional Deviant Disorder for the SDQ conduct scale, Attention-Deficit/Hyperactivity Disorder for the SDQ hyperactivity scale and autism spectrum disorder for the SDQ social problems and prosocial behavior scales. For the SDQ total scale, the total clinical sample was used.

the SDQ total difficulties scale can be used to indicate whether an adolescent likely belongs to the clinical population or not. Note that when taking into account the adolescents' gender, the adolescent-reported total difficulties scale was found to distinguish sufficiently well for female adolescents but not for male, indicating that the adolescent-reported total difficulties scale can be used to screen for psychosocial problems among female adolescents and that the same scale of the parent-reported version is useful for both males and females. For all other SDQ scales, potentially useful for screening for specific types of disorders, no gender differences were found.

Regarding the specific SDQ difficulties and strength scales, *both* SDQ versions' emotional problems, conduct problems, and hyperactivity/inattention scales appeared sufficiently capable of distinguishing between the community sample and adolescents diagnosed with an anxiety/mood disorder, CD/ODD, and ADHD, respectively. We have not been able to compare our findings with previous research as, to the best of our knowledge, the criterion validity of the SDQ difficulties scales, other than the aforementioned total difficulties scale, has not been investigated previously. Note that perfect distinction between community and clinical (sub)populations cannot be expected as (a) in the community population some undetected psychiatric disorders can be expected to be prevalent and (b) adolescents in the clinical population do not only receive *DSM-IV* diagnoses in one of the four categories that are content-wise corresponding to the SDQ scales. Moreover, the results may be biased to some extent as it is likely that adolescents with worrisome but minor psychosocial problems are underrepresented in our clinical sample as they may not (yet) be referred to mental health care.

Overall, our findings regarding the criterion validity of the SDQ difficulties scales suggest that they can be used to

screen for the problems related to anxiety/mood disorder, CD/ODD, and ADHD among community adolescent populations. Keep in mind that the SDQ was not developed for diagnostic purposes; after the SDQ is used to provide a preliminary indication of potential problems at hand, thorough assessment by clinicians is needed.

For the SDQ *parent* version the social problems scale was found to sufficiently distinguish between the community sample and the clinical sample diagnosed with ASD. In contrast, the parent-reported prosocial behavior scale and both the adolescent self-reported social problems and prosocial behavior scales appear insufficiently useful for discriminating between community adolescents and adolescents diagnosed with ASD. In other words, the parent appears to be a better informant for ASD than the adolescent, whereby the parent-reported SDQ social problems scale is a useful indicator and the prosocial behavior scale is not.

Limitations

The preceding discussion of the outcomes of our study implies several strengths. Besides advancing previous research in multiple respects, however, the current study is prone to some potential limitations. First, the community sample data used in this study was gathered in two waves, approximately 7 years apart. Moreover, the community sample is not fully representative of the population of Dutch adolescents as adolescents with a mother born in the Netherlands (as opposed to a mother born in another country), adolescents with a mother with a medium educational level (as opposed to low or high), and adolescents living in the East and West of the Netherlands were slightly overrepresented in the community sample. Additionally, the sampling strategies resulted in overrepresentation of 13- and

14-year-olds. By handling these data as being representative of the Dutch adolescent community population, we assume that validity aspects do not change over time and do not depend on characteristics such as ethnicity and age. Though we consider these assumptions to be reasonable, we cannot rule out that the small deviations from the population distribution have resulted in slightly biased results.

The second limitation follows from the fact that our community sample contained missing data at two levels: questionnaire level and item level. First, regarding missing data at questionnaire level, all adolescents had data available of at least one SDQ version. For a subset of these adolescents, CBCL/YSR and/or IDS-2 data were available. The missingness of the second SDQ version and the CBCL/YSR questionnaires may not be random, but considering the large numbers of questionnaires that are available to us, we expect the outcomes of this study to be minimally affected. The missingness of IDS-2 questionnaires definitely is not random as only a subsample of the adolescents with at least one SDQ version available was approached to complete the IDS-2. The adolescents in this subsample showed a relatively low average IQ score and are thus IQ-wise not representative of the population of Dutch adolescents. As we do not know whether the way in which the SDQ measures psychosocial functioning differs across lower and average IQ's, this too, may have biased our results to some extent. Second, regarding missing data at item level and taking into consideration the relatively small numbers of missing SDQ, CBCL/YSR, and IDS-2 data, we expect the potential bias in our outcomes to be minimal.

Conclusion

The SDQ is widely used to screen for psychosocial problems in community settings. In this study, we found some support for the SDQ's intended scale structure (emotional problems, conduct problems, hyperactivity/inattention, social problems, and prosocial behavior). However, both SDQ versions had some questionable indicators, unintended subfactors, and insufficient scale reliabilities, suggesting that the SDQ's presumed scale structure is not fully tenable among adolescents in a screening setting. In contrast, the results also suggest that the SDQ scales, using CBCL/YSR and IDS-2 scales as criteria, measure the intended types of difficulties and do not appear to unintendedly measure other aspects of behavior or intelligence. Moreover, the results indicate that both adolescent- and parent-rated SDQ scores can be used to distinguish adolescents likely belonging to the clinical population from other adolescents and that individual scales from both SDQ versions can be used to identify adolescents with specific types of disorders (parent and adolescent: anxiety/mood disorder, CD/ODD, ADHD; only parent: ASD). Evidence regarding the SDQ's scale structure warrants some caution for the use of the scales in their current form. However, the evidence regarding the various validity aspects are mostly supportive for the continued use of the SDQ adolescent and parent versions as currently used for screening in routine well-child care practice among adolescents.

Supplemental Material

Supplemental material for this article is available online.

ORCID iDs

Jorien Vugteveen https://orcid.org/0000-0002-8098-4120
Annelies de Bildt https://orcid.org/0000-0002-4196-2404

References

Achenbach, T. M. (1991a). *Manual for the child behavior checklist/4-18 and 1991 profile*. Burlington: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1991b). *Manual for the youth self-report and 1991 profile*. Burlington: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (2014). *DSM-oriented guide for the Achenbach System of Empirically Based Assessment (ASEBA)*. Burlington: University of Vermont, Research Center for Children, Youth, and Families.

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*, 397-438.

Asparouhov, T., Muthén, B., & Morin, A. J. (2015). Bayesian structural equation modeling with cross-loadings and residual covariances: Comments on Stromeyer et al. *Journal of Management*, *41*, 1561-1577.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238-246.

Bøe, T., Hysing, M., Skogen, J. C., & Breivik, K. (2016). The Strengths and Difficulties Questionnaire (SDQ): Factor structure and gender equivalence in Norwegian adolescents. *PLoS ONE*, *11*, e0152202.

Chen, W. J., Faraone, S. V., Biederman, J., & Tsuang, M. T. (1994). Diagnostic accuracy of the Child Behavior Checklist scales for attention-deficit hyperactivity disorder: A receiver-operating characteristic analysis. *Journal of Consulting and Clinical Psychology*, *62*, 1017-1025.

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

Evers, A., Sijtsma, K., Lucassen, W., & Meijer, R. R. (2010). The Dutch review process for evaluating the quality of psychological tests: History, procedure, and results. *International Journal of Testing*, *10*, 295-317.

Fergusson, D. M., Horwood, L. J., & Lynskey, M. T. (1993). Prevalence and comorbidity of *DSM-III-R* diagnoses in a birth cohort of 15 year olds. *Journal of the American Academy of Child & Adolescent Psychiatry*, *32*, 1127-1134.

Garrido, L. E., Barrada, J. R., Aguasvivas, J. A., Martínez-Molina, A., Arias, V. B., Golino, H. F., . . . Rojo-Moreno, L. (2020). Is small still beautiful for the Strengths and Difficulties Questionnaire? Novel findings using exploratory structural equation modeling. *Assessment*, *27*(6), 1349-1367. doi:10.1177/1073191118780461

Giannakopoulos, G., Tzavara, C., Dimitrakaki, C., Kolaitis, G., Rotsika, V., & Tountas, Y. (2009). The factor structure of the Strengths and Difficulties Questionnaire (SDQ) in Greek adolescents. *Annals of General Psychiatry*, *8*, 20. doi:10.1186/1744-859X-8-20

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 & Adolescent Psychiatry*, *40*, 1337-1345.

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

Grob, A., Hagmann-von Arx, P., Ruiter, S., Timmerman, M. E., & Visser, L. (2018). *IDS-2: Intelligentie- en ontwikkelingsschalen voor kinderen en jongeren* [IDS-2: Intelligence and development scales for children and adolescents]. Amsterdam, Netherlands: Hogrefe.

Grob, A., Meyer, C., & Hagmann-von Arx, P. (2018). *Intelligence and development scales-2*. Bern, Switzerland: Hogrefe.

He, J., Burstein, M., Schmitz, A., & Merikangas, K. R. (2013). The Strengths and Difficulties Questionnaire (SDQ): The factor structure and scale validation in U.S. adolescents. *Journal of Abnormal Child Psychology*, *41*, 583-595.

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

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2018). *semTools: Useful tools for structural equation modeling* (R Package Version 0.5-1). Retrieved from https://github.com/simsem/semTools/wiki

Koskelainen, M., Sourander, A., & Vauras, M. (2001). Self-reported strengths and difficulties in a community sample of Finnish adolescents. *European Child & Adolescent Psychiatry*, *10*, 180-185.

Lundh, L., Wångby-Lundh, M., & Bjärehed, J. (2008). Self-reported emotional and behavioral problems in Swedish 14- to 15-year-old adolescents: A study with the self-report version of the Strengths and Difficulties Questionnaire. *Scandinavian Journal of Psychology*, *49*, 523-532.

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

Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual review of Clinical Psychology*, *10*, 85-110.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*, 115-132.

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.

Nakamura, B. J., Ebesutani, C., Bernstein, A., & Chorpita, B. F. (2009). A psychometric analysis of the child behavior checklist DSM-oriented scales. *Journal of Psychopathology and Behavioral Assessment*, *31*, 178-189.

Ormel, J., Raven, D., van Oort, F., Hartman, C., Reijneveld, S., Veenstra, R., . . . Oldehinkel, A. (2015). Mental health in Dutch adolescents: A TRAILS report on prevalence, severity, age of onset, continuity and co-morbidity of DSM disorders. *Psychological Medicine*, *45*, 345-360.

Ortuño-Sierra, J., Fonseca-Pedrero, E., Paino, M., Sastre, i, Riba, S., & Muñiz, J. (2015). Screening mental health problems during adolescence: Psychometric properties of the Spanish version of the Strengths and Difficulties Questionnaire. *Journal of Adolescence*, *38*, 49-56.

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Richter, J., Sagatun, Å., Heyerdahl, S., Oppedal, B., & Røysamb, E. (2011). The Strengths and Difficulties Questionnaire (SDQ)–Self-Report: An analysis of its structure in a multiethnic urban adolescent sample. *Journal of Child Psychology and Psychiatry*, *52*, 1002-1011.

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 Bioinformatics*, *12*, 77. doi:10.1186/1471-2105-12-77

Rønning, J. A., Handegaard, B. H., Sourander, A., & Mørch, W. (2004). The Strengths and Difficulties Self-Report

Questionnaire as a screening instrument in Norwegian community samples. *European Child & Adolescent Psychiatry*, *13*, 73-82.

Ruchkin, V., Koposov, R., & Schwab-Stone, M. (2007). The Strength and Difficulties Questionnaire: Scale validation with Russian adolescents. *Journal of Clinical Psychology*, *63*, 861-869.

Saris, W. E., Satorra, A., & van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling: A Multidisciplinary Journal*, *16*, 561-582. doi:10.1080/10705510903203433

Statistics Netherlands. (2015). *Statline*. Retrieved from https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37296ned/table?ts=152209294

Steiger, J. H. (1980, MAY). *Statistically based tests for the number of common factors*. Paper presented at the Annual Meeting of the Psychometric Society, Iowa City, IA.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*, 1-10.

Van der Ark, L. A. (2007). Mokken scale analysis in R. *Journal of Statistical Software*, *20*(11), 1-19.

Van de Looij-Jansen, P. M., Goedhart, A. W., de Wilde, E. J., & Treffers, P. D. A. (2011). Confirmatory factor analysis and factorial invariance analysis of the adolescent self-report Strengths and Difficulties Questionnaire: How important are method effects and minor factors? *British Journal of Clinical Psychology*, *50*, 127-144.

Van Ginkel, J. R., Ark, L. A., & Sijtsma, K. (2007). Multiple imputation for item scores when test data are factorially complex. *British Journal of Mathematical and Statistical Psychology*, *60*, 315-337.

Van Lang, N. D., Ferdinand, R. F., Oldehinkel, A. J., Ormel, J., & Verhulst, F. C. (2005). Concurrent validity of the DSM-IV scales affective problems and anxiety problems of the youth self-report. *Behaviour Research and Therapy*, *43*, 1485-1494.

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.

Van Widenfelt, B. M., Goedhart, A. W., Treffers, P. D., & Goodman, R. (2003). Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *European Child & Adolescent Psychiatry*, *12*, 281-289.

Verhulst, F. C., Van der Ende, J., & Koot, H. M. (1996). *Dutch manual for the CBCL/4–18*. Rotterdam, Netherlands: Afdeling Kinder-En Jeugdpsychiatrie.

Verhulst, F. C., Van der Ende, J., & Koot, H. M. (1997). *Dutch manual for the youth self-report (YSR)*. Rotterdam, Netherlands: Afdeling Kinder-En Jeugdpsychiatry.

Vogels, A. G. C., Siebelink, B. M., Theunissen, M. H. C., Wolff, M. S. d., & Reijneveld, S. A. (2011). *Vergelijking van de KIVPA en de SDQ als signaleringsinstrument voor problemen bij adolescenten in de jeugdgezondheidszorg* [Comparison of the KIVPA and SDQ as used for screening among adolescents in youth healthcare]. Leiden, Netherlands: TNO.

Vugteveen, J., de Bildt, A., Hartman, C. A., & Timmerman, M. (2018). Using the Dutch multi-informant Strengths and Difficulties Questionnaire (SDQ) to predict adolescent psychiatric diagnoses. *European Child & Adolescent Psychiatry*, *27*, 1347-1359. doi:10.1007/s00787-018-1127-y

Vugteveen, J., de Bildt, A., Serra, M., de Wolff, M., & Timmerman, M. (2020). Psychometric properties of the Dutch Strengths and Difficulties Questionnaire (SDQ) in adolescent community and clinical populations. *Assessment*, *27*(7), 1476-1489. doi:10.1177/1073191118804082

Yang, Y., & Green, S. B. (2015). Evaluation of structural equation modeling estimates of reliability for scales with ordered categorical items. *Methodology*, *11*, 23-34.

Supplementary Materials

Paper title: Validity aspects of the Strengths and Difficulties Questionnaire (SDQ) adolescent self-report and parent-report versions among Dutch adolescents

Authors: Jorien Vugteveen, Annelies de Bildt, Meinou Theunissen, Menno Reijneveld, and Marieke Timmerman

Table S1 Available questionnaires within the community sample (n = 962)

		Achenbach questionnaires					IDS-2
		YSR	CBCL	YSR & CBCL	None	Total	
SDQ version	Adolescent	361	0	33	43	437	81
	Parent	0	18	10	17	45	13
	Adolescent and parent	31	9	425	15	480	126
	Total	392[a]	27[b]	468[c]	75	962	220

SDQ: Strengths and Difficulties Questionnaire; YSR: Youth Self Report; CBCL: Child Behavior Checklist.
IDS-2: Intelligence Development Scales 2.
[a] incl. 9 adolescents who provided too few item scores to calculate at least one YSR scale score.
[b] incl. 5 parents who provided too few item scores to calculate at least one CBCL scale score.
[c] incl. 1 pair of parent and adolescent who provided to few item scores to calculate at least one CBCL of YSR scale.

Table S2 Prevalence of DSM-IV diagnoses and comorbid DSM-IV diagnoses in the clinical

sample (n = 4,053)

| DSM category[a] | N[b] | Comorbid with ... | | | |
		ADHD	CD/ODD	Anxiety/mood disorder	ASD
ADHD[c]	913	-	.18	.14	.16
Anxiety/Mood disorder[c]	1,372	.09	.03	-	.09
ASD[c]	719	.20	.04	.18	-
CD/ODD[c]	391	.42	-	.09	.08

[a] *ADHD: Attention-Deficit/Hyperactivity Disorder, ASD: Autism Spectrum Disorder, CD/ODD: Conduct/Oppositional Defiant Disorder*
[b] *The numbers in this column add up to more than 2,812 (sample size) due to comorbidity*
[c] *The proportion of adolescents within each DSM category (row), also diagnosed with any of the other disorders*

Table S3 Goodness-of-fit statistics and correlated errors of the CFA and ESEM models for the SDQ adolescent and parent versions in the community sample

Model	χ^2	*df*	*p*-value	RMSEA	RMSEA 90% CI	CFI	TLI	SEPC	Θ1	Θ2	Θ3	Θ4
SDQ adolescent version												
CFA-5F	772.988	265	<.001	.046	[.042 - .049]	.896	.883					
CFA-6F	525.249	255	<.001	.034	[.030 - .038]	.945	.935					
	493.887	254	<.001	.032	[.028 - .036]	.951	.942	Θ1 = Θ$_{25,15}$ = .488	.392			
	484.004	253	<.001	.032	[.027 - .036]	.953	.944	Θ2 = Θ$_{24,16}$ = .327	.392	.277		
Final	464.121	252	<.001	.030	[.026 - .035]	.957	.948	Θ3 = Θ$_{20,9}$ = .313	.391	.277	.278	
ESEM-5F	304.576	185	<.001	.027	[.021 - .032]	.976	.960					
	272.212	184	<.001	.023	[.017 - .028]	.982	.971	Θ1 = Θ$_{10,2}$ = .930	.436			
	259.109	183	<.001	.021	[.015 - .027]	.984	.975	Θ2 = Θ$_{19,14}$ = .364	.438	.318		
Final	246.660	182	.001	.020	[.013 - .026]	.987	.978	Θ3 = Θ$_{13,3}$ = .274	.440	.317	.253	
SDQ parent version												
CFA-5F	576.368	265	<.001	.047	[.042 - .053]	.926	.916					
	542.354	264	<.001	.045	[.039 - .050]	.934	.925	Θ1 = Θ$_{10,2}$ = .784	.539			
	529.366	263	<.001	.044	[.038 - .049]	.937	.928	Θ2 = Θ$_{13,8}$ = .721	.539	.478		
	518.685	262	<.001	.043	[.038 - .049]	.939	.930	Θ3 = Θ$_{22,18}$ = .542	.539	.477	.470	
Final	502.226	261	<.001	.042	[.036 - .047]	.943	.934	Θ4 = Θ$_{20,9}$ = .433	.539	.477	.470	.359
ESEM-5F	274.950	185	<.001	.030	[.023 - .038]	.979	.965					
	250.658	184	<.001	.026	[.017 - .034]	.984	.974	Θ1 = Θ$_{25,15}$ = 1.733	.569			
Final	239.155	183	.003	.024	[.014 - .032]	.987	.978	Θ2 = Θ$_{24,16}$ = .658	.547	.417		

Notes. For the SDQ adolescent version: n = 917; for the SDQ parent version: n = 525.
χ^2: chi square value*; df:* degrees of freedom; RMSEA: root mean square error of approximation; CFI: comparative fit index; TLI: Tucker-Lewis index; SEPC: standardized expected parameter change; Θ: correlated residuals

Table S4 Standardized parameter estimates of the final CFA and ESEM models with correlated errors for the SDQ adolescent version

Item/ factor	CFA six-factor model						ESEM five-factor model				
	ES	CP	HP	SP	PB	PCM	ES	CP	HP	SP	PB
3	**.49**						**.38**	.12	.10	-.05	.03
8	**.74**						**.67**	.13	.02	.07	.16
13	**.80**						**.62**	.04	.08	.19	.09
16	**.59**						**.70**	-.07	-.02	.01	-.10
24	**.73**						**.76**	.18	.06	.15	-.06
5		**.78**					.30	**.61**	.08	.03	.02
7		**.03**				.63	-.01	**.27**	.18	-.23	-.31
12		**.61**					-.19	**.58**	.10	.13	-.06
18		**.67**					-.17	**.57**	.15	.29	.03
22		**.62**					.05	**.53**	.01	.09	-.09
2			**.82**				-.02	.12	**.66**	-.02	.08
10			**.77**				.18	.17	**.49**	-.05	.09
15			**.73**				-.04	-.14	**.99**	.10	.11
21			**.34**			.41	.11	.29	**.33**	-.24	-.19
25			**.38**			.33	.004	-.10	**.65**	-.05	-.21
6				**.64**			.21	.05	-.07	**.54**	-.14
11				**.39**		.30	.01	.12	.03	**.31**	-.25
14				**.58**		.36	.13	.22	-.01	**.31**	-.27
19				**.74**			.15	.21	.06	**.48**	.09
23				**.55**			.08	.04	-.03	**.56**	.01
1					**.78**		.21	-.37	-.01	-.12	**.51**
4					**.45**		-.07	-.05	.08	-.14	**.42**
9					**.55**		.06	.23	-.12	-.21	**.72**
17					**.64**		-.06	-.17	-.01	-.02	**.50**
20					**.48**		-.05	.05	-.01	.12	**.68**

	Factor correlations										
	ES	CP	HP	SP	PB	PCM	ES	CP	HP	SP	PB
ES	1.00	.34	.38	.60	-.01	-.03	1.00	.12	.27	.38	.02
CP		1.00	.53	.50	-.55	.44		1.00	.43	.27	-.33
HP			1.00	.17	-.21	.36			1.00	.09	-.24
SP				1.00	-.30	-.07				1.00	-.33
PB					1.00	-.73					1.00
PCM						1.00					

ESEM = exploratory structural equation modelling, CFA = confirmatory factor analysis, ES = emotional symptoms, CP = conduct problems, HP = hyperactivity/attention problems, SP = social problems, PB = prosocial behaviour, PCM = positive construal method. Per item, its loading on its intended factor is printed in bold.

Table S5 Standardized parameter estimates of the CFA and ESEM models with correlated errors for the SDQ parent version

Item/ factor	CFA five-factor model					ESEM five-factor model				
	ES	CP	HP	SP	PB	ES	CP	HP	SP	PB
3	**.34**					**.49**	.05	-.01	-.21	.01
8	**.76**					**.91**	.01	-.08	.09	.12
13	**.69**					**.82**	-.16	.02	.09	.01
16	**.80**					**.50**	.13	.05	.19	.01
24	**.80**					**.45**	.21	-.01	.26	.01
5		**.60**				.38	**.17**	.19	-.10	-.10
7		**.55**				.09	**.30**	.18	-.20	-.35
12		**.40**				.13	**.22**	.21	.07	.11
18		**.65**				.01	**.57**	.25	-.05	-.13
22		**.40**				-.09	**1.06**	-.20	.15	.15
2			**.68**			-.18	-.02	**.93**	.14	.09
10			**.65**			-.07	-.08	**.88**	.20	.18
15			**.88**			.06	.13	**.66**	.02	.03
21			**.63**			.07	.09	**.54**	-.23	-.25
25			**.84**			.15	.07	**.58**	-.11	-.13
6				**.53**		.12	.01	-.13	**.47**	-.22
11				**.63**		.002	-.02	.10	**.59**	-.18
14				**.75**		.15	.04	.08	**.42**	-.34
19				**.80**		.28	.11	.24	**.51**	.09
23				**.66**		.08	.06	.01	**.65**	-.09
1					**.91**	-.01	-.19	-.15	-.12	**.64**
4					**.78**	-.09	-.08	.13	-.22	**.67**
9					**.68**	.09	.09	-.02	-.13	**.80**
17					**.62**	-.04	.26	.02	-.22	**.64**
20					**.53**	.15	-.15	.03	.13	**.77**

Factor correlations

	ES	CP	HP	SP	PB	ES	CP	HP	SP	PB
ES	1.00	.57	.42	.71	-.24	1.00	.30	.39	.43	-.17
CP		1.00	.74	.49	-.60		1.00	.39	.14	-.20
HP			1.00	.39	-.31			1.00	.15	-.22
SP				1.00	-.59				1.00	-.23
PB					1.00					1.00

ESEM = exploratory structural equation modelling, CFA = confirmatory factor analysis, ES = emotional symptoms, CP = conduct problems, HP = hyperactivity/attention problems, SP = social problems, PB = prosocial behaviour, PCM = positive construal method. Per item, its loading on its intended factor is printed in bold.

SDQ self-report version
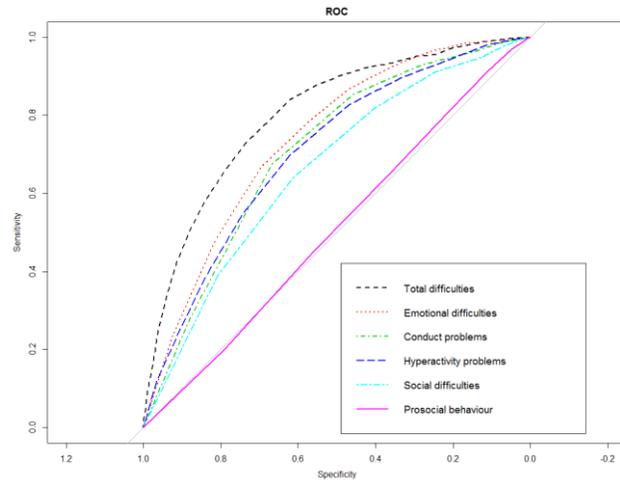
Fig. S1 Using SDQ scales to distinguish between the community and clinical samples
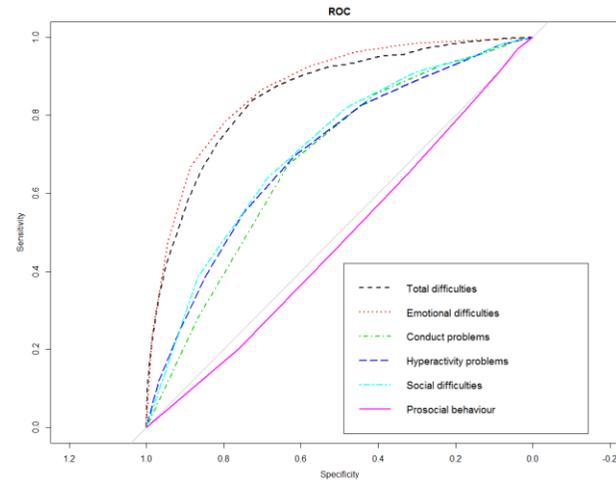
Fig. S2 Using SDQ scales to distinguish between the community sample and clinical sample diagnosed with Anxiety/mood disorder
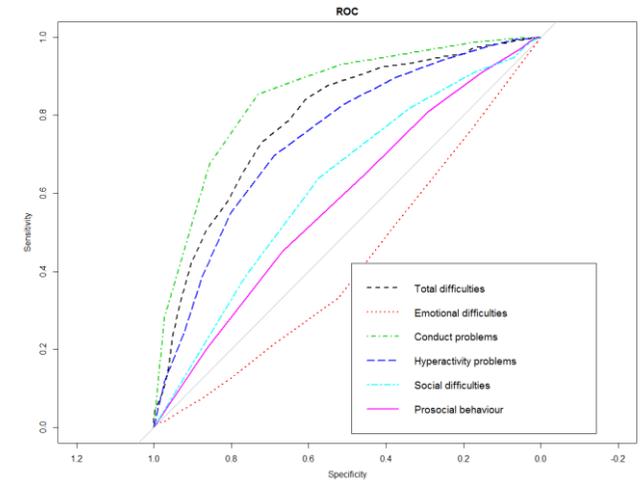
Fig. S3 Using SDQ scales to distinguish between the community sample and clinical sample diagnosed with CD/ODD
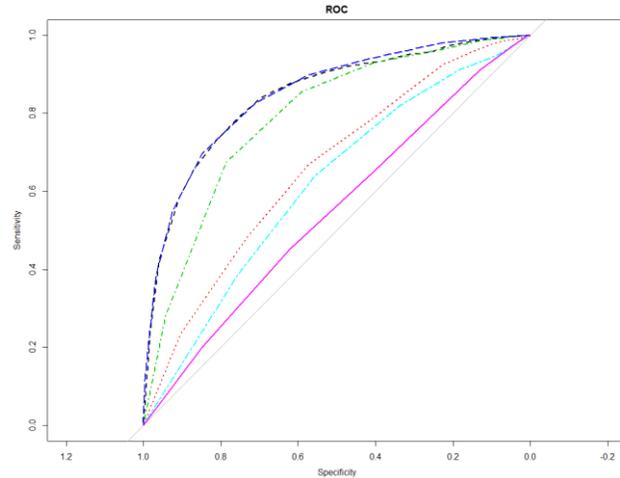

Fig. S4 Using SDQ scales to distinguish between the community sample and clinical sample diagnosed with ADHD
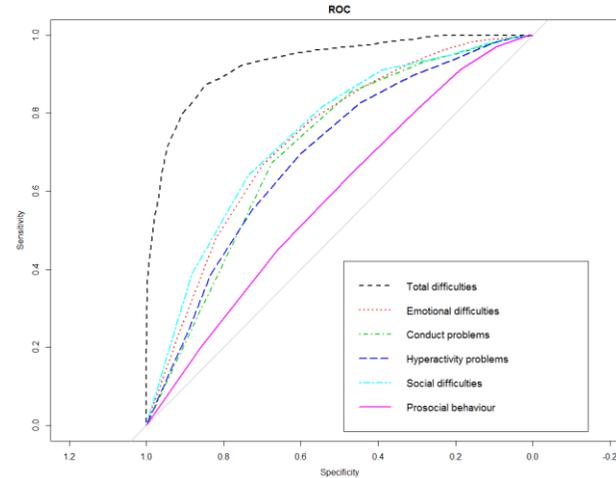
Fig. S5 Using SDQ scales to distinguish between the community sample and clinical sample diagnosed with ASD

SDQ parent-report version
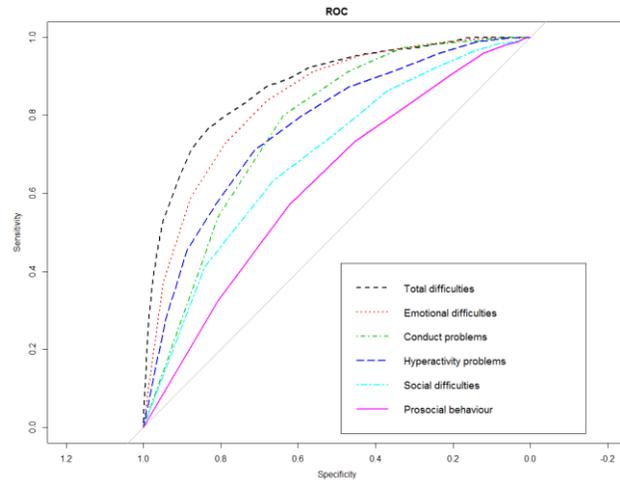
Fig. S6 Using SDQ scales to distinguish between the community and clinical samples
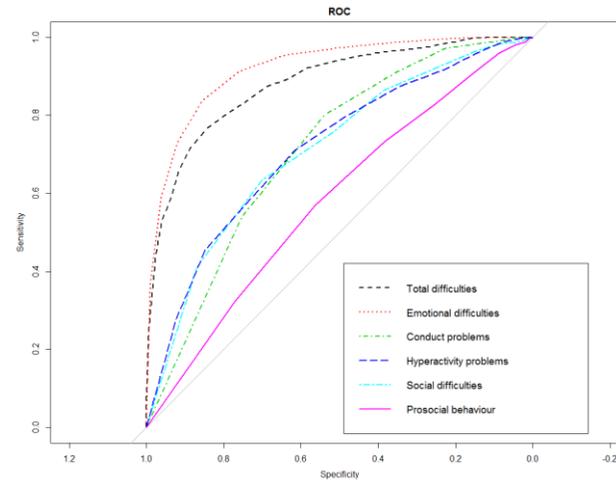
Fig. S7 Using SDQ scales to distinguish between the community sample and clinical sample diagnosed with Anxiety/mood disorder
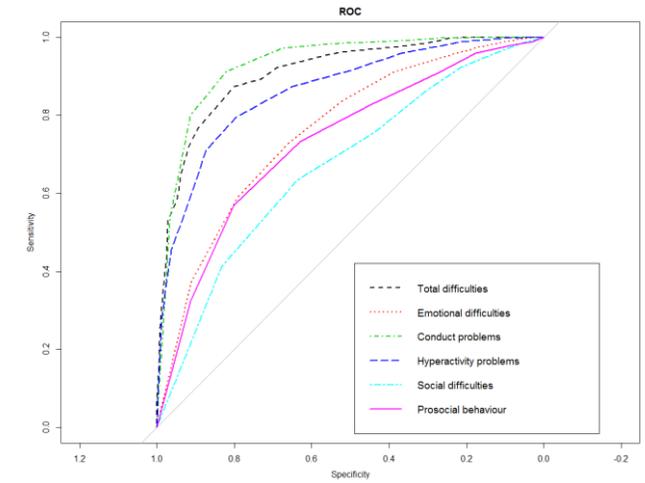
Fig. S8 Using SDQ scales to distinguish between the community sample and clinical sample diagnosed with CD/ODD
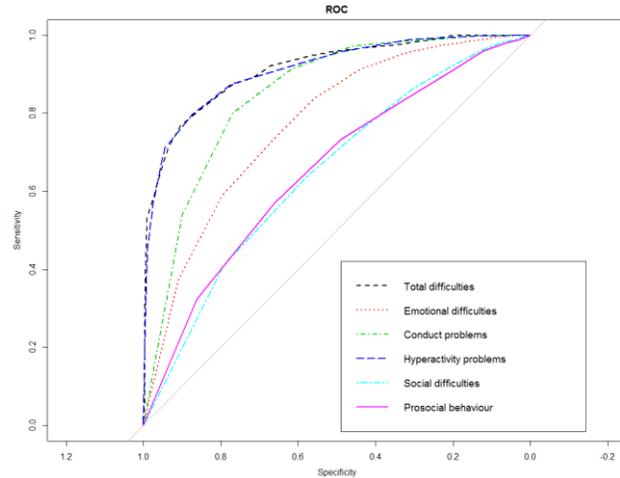
Fig. S9 Using SDQ scales to distinguish between the community sample and clinical sample diagnosed with ADHD
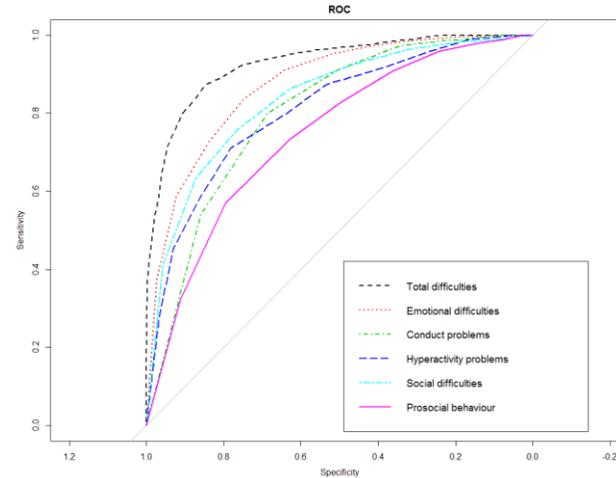
Fig. S10 Using SDQ scales to distinguish between the community sample and clinical sample diagnosed with ASD

Per setting and gender, Table S6 provides mean scale scores and cronbach's alpha coefficients. For the SDQ *adolescent version*, gender differences were found for the prosocial behaviour scale and the emotional, conduct and social difficulties scales in the community setting. In the clinical setting, gender differences were found for the same scales and additionally for the total difficulties scale. For the *parent version*, gender differences were found for the prosocial behaviour scale and the hyperactivity, social and total difficulties scales in the community setting. For the clinical setting, differences were found for all scales except the social difficulties scales. As SDQ scores differ across gender, it is not unlikely that gender differences also exist in the SDQ scales' ability to distinguish between groups. The AUC values per SDQ version and gender are provided in Table S4. Graphical representations are provides as well (Figures S11 to S30). Considering AUC values ≥ .80 as indicating sufficient ability to distinguish between samples, the main gender difference is found for the SDQ adolescent version's total difficulties scale. For females, this scale is sufficiently accurate at distinguishing between the community and clinical samples, for males it is not.

Table S6 Per SDQ version (adolescent, parent) and per setting (community, clinical): Mean scale scores, standard deviations and Cronbach's Alpha for female and male adolescents

		Community setting				Clinical setting			
		Female[a]		Male[b]		Female[a]		Male[b]	
	SDQ scale	α[c]	*M (SD)*	α[c]	*M (SD)*	α[c]	*M (SD)*	α[c]	*M (SD)*
SDQ adolescent version									
	Total[c]	.78	8.1 (4.8)	.75	8.1 (4.7)	.77	15.6 (5.9)	.76	13.3 (5.7)
	Emotional	.71	2.6 (2.2)	.56	1.6 (1.6)	.73	5.5 (2.6)	.70	3.1 (2.4)
	Conduct	.45	1.1 (1.1)	.51	1.6 (1.5)	.57	2.4 (1.8)	.59	2.8 (1.9)
	Hyper	.77	3.2 (2.3)	.71	3.5 (2.3)	.75	5.2 (2.6)	.76	5.3 (2.6)
	Social	.53	1.2 (1.4)	.53	1.5 (1.6)	.52	2.4 (1.9)	.55	2.1 (1.8)
	Prosocial	.53	8.4 (1.4)	.59	7.5 (1.7)	.62	8.2 (1.7)	.64	7.5 (1.9)
SDQ parent version									
		Female[a]		Male[b]		Female[a]		Male[b]	
	SDQ scale	α[c]	*M (SD)*	α[c]	*M (SD)*	α[c]	*M (SD)*	α[c]	*M (SD)*
	Total[d]	.78	5.5 (4.6)	.81	7.4 (5.3)	.79	15.6 (6.5)	.79	16.2 (6.5)
	Emotional	.67	1.7 (1.9)	.68	1.6 (1.9)	.66	5.8 (2.6)	.67	4.2 (2.7)
	Conduct	.34	0.7 (1.0)	.55	0.9 (1.3)	.72	2.5 (2.2)	.74	3.1 (2.5)
	Hyper	.76	3.0 (2.5)	.78	1.9 (2.1)	.75	4.5 (2.7)	.74	5.9 (2.6)
	Social	.61	1.1 (1.6)	.65	1.9 (2.1)	.64	2.9 (2.2)	.68	3.0 (2.4)
	Prosocial	.72	8.6 (1.7)	.71	8.0 (1.9)	.75	7.6 (2.2)	.73	7.0 (2.2)

SDQ: Strengths and Difficulties questionnaire; α: Cronbach's index of internal consistency (alpha)
[a] Adolescent version community setting: N = 457 (female), N = 442 (male); clinical setting: N = 2,002 (female), N = 1.792 (male).
[b] Parent version community setting: N = 252 (female), N = 240 (male); clinical setting: N = 1,898 (female), N = 1.755 (male).
[c] For the SDQ adolescent version, t-tests for comparing means revealed gender differences for all scales with the exception of the SDQ total difficulties scale (community setting) and the hyperactivity scale (both settings).
[d] For the SDQ parent version, t-tests for comparing means revealed gender differences for all scales with the exception of the emotional and conduct difficulties scales (community setting) and the social problems scale (clinical setting). reported prosocial behaviour scale.

Table S7 Per SDQ version and gender, the SDQ scales' abilities to distinguish between community and clinical (sub)samples

SDQ scale	SDQ adolescent version					
	Female			Male		
	Comm. N	*Clin. N[a]*	*AUC (SE)*	*Comm. N*	*Clin. N*	*AUC (SE)*
Total	461	2002	.84 (.01)	450	1792	.76 (.01)[b]
Emotional	461	934	.87 (.01)	450	385	.84 (.01)
Conduct	461	101	.90 (.02)	450	256	.81 (.02)
Hyper	461	284	.89 (.01)	450	583	.83 (.01)
Social	461	231	.79 (.02)	450	429	.71 (.02)
Prosocial	461	231	.62 (.02)	450	429	.54 (.02)
SDQ scale	SDQ parent version					
	Female			Male		
	Comm. N	*Clin. N[a]*	*AUC (SE)*	*Comm. N*	*Clin. N*	*AUC (SE)*
Total	271	1898	.90 (.01)	251	1755	.85 (.01)[c]
Emotional	271	850	.93 (.01)	251	359	.89 (.01)
Conduct	271	103	.91 (.02)	251	237	.93 (.01)
Hyper	271	279	.92 (.01)	251	570	.89 (.01)
Social	271	233	.87 (.02)	251	431	.80 (.02)
Prosocial	271	233	.75 (.02)	251	431	.73 (.02)

SDQ: Strengths and Difficulties Questionnaire; Comm.: Community sample; Clin.: Clinical (sub)sample; AUC: Area Under the Curve
[a] Per SDQ scale, the clinical subsamples consisted of adolescent with a DSM-IV diagnosis content-wise matching the SDQ scale: Anxiety/Mood disorder for the SDQ emotional scale, Conduct / Oppositional Deviant Disorder for the SDQ conduct scale, Attention-Deficit/Hyperactivity Disorder for the SDQ hyperactivity scale and Autism Spectrum Disorder for the SDQ social problems and prosocial behaviour scales. For the SDQ total scale, the total clinical sample was used.
[b] For the SDQ adolescent version, row wise comparison of the ROC values revealed significant gender differences for all scales except the emotional difficulties scale
[c] For the SDQ parent version, row wise comparison of the ROC values revealed significant gender differences for the total, emotional and social difficulties scales.

SDQ self-report version

Using SDQ scales to distinguish between the community and clinical samples

Using SDQ scales to distinguish between the community sample and clinical sample diagnosed with Anxiety/mood disorder

Using SDQ scales to distinguish between the community sample and clinical sample diagnosed with CD/ODD
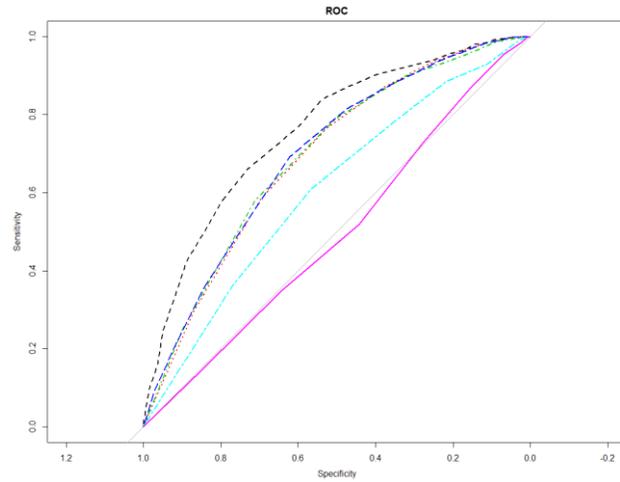
Fig. S11 female adolescents

*Fig. S13* female adolescents

Fig. S15 female adolescents

Fig. S12 male adolescents
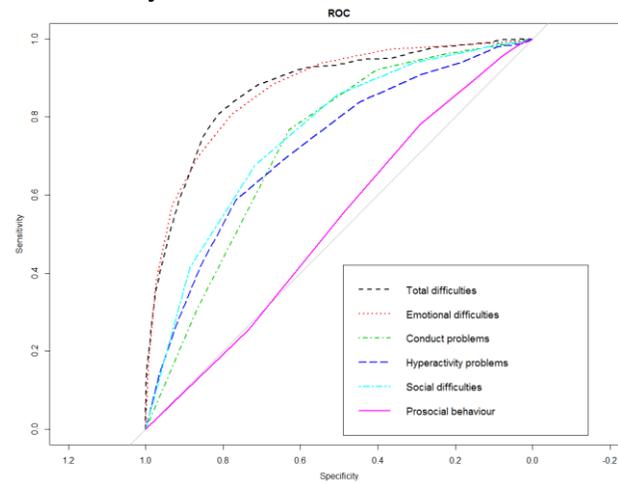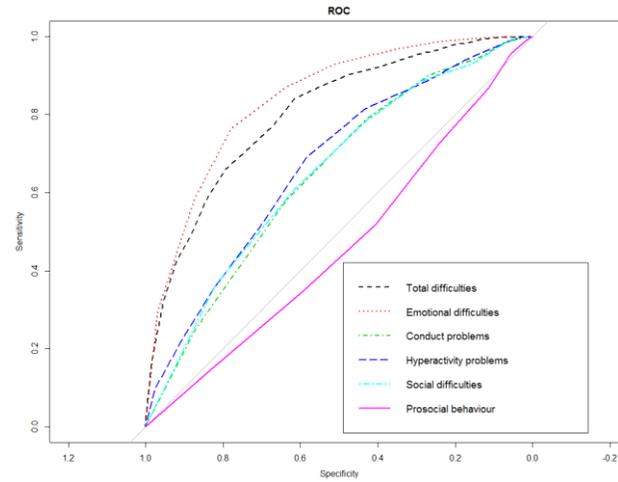
Fig. S14 male adolescents
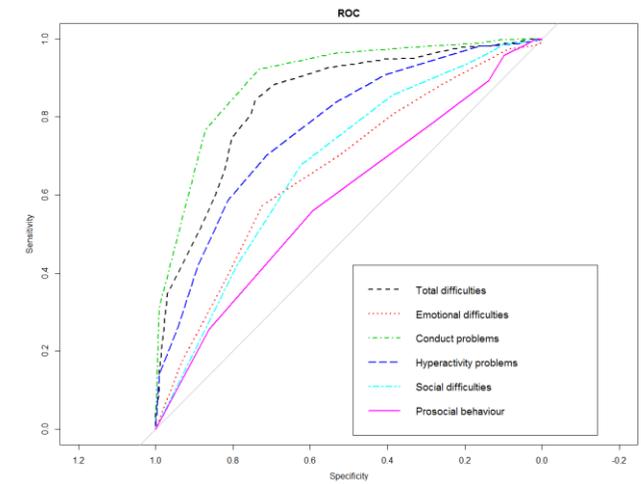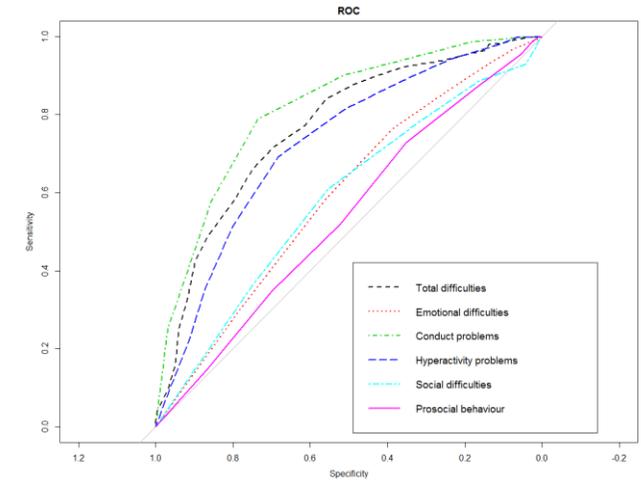
Fig. S16 male adolescents

Using SDQ scales to distinguish between the
community sample and clinical sample diagnosed
with ADHD
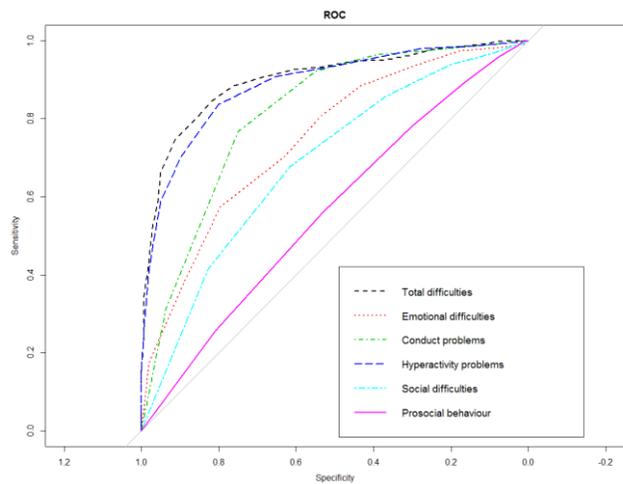
Fig. S17 female adolescents

Using SDQ scales to distinguish between the
community sample and clinical sample diagnosed
with ASD
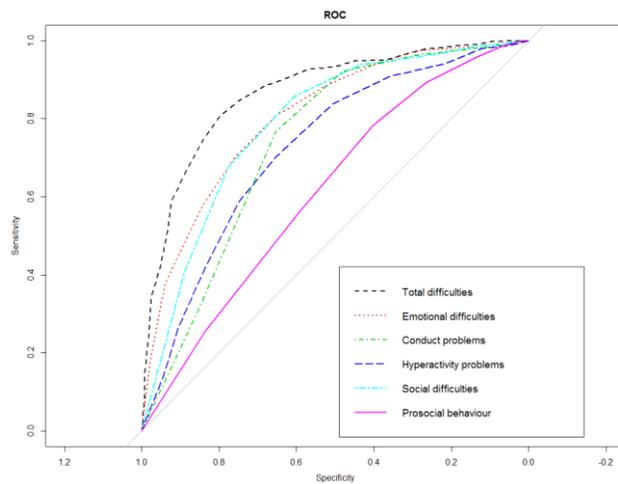
Fig. S19 female adolescents
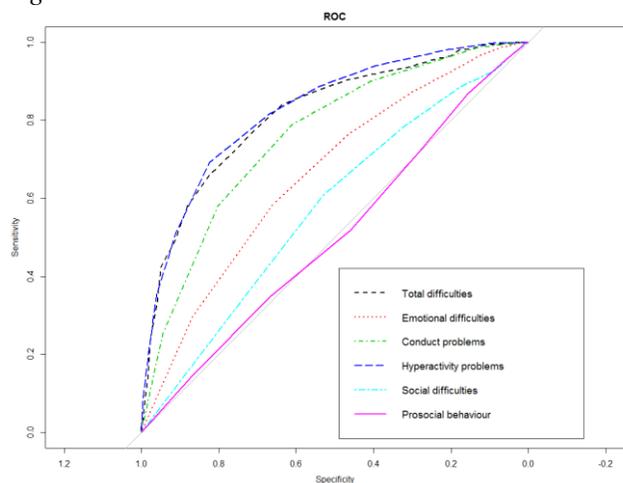
Fig. S18 male adolescents
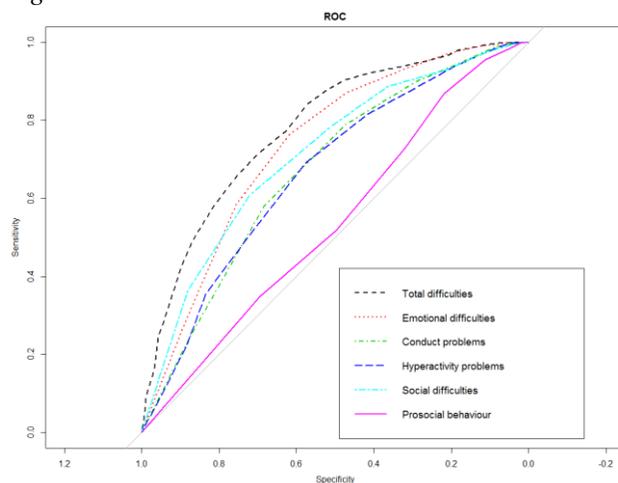
Fig. S20 male adolescents

SDQ parent-report version

Using SDQ scales to distinguish between the community and clinical samples

Using SDQ scales to distinguish between the community sample and clinical sample diagnosed with Anxiety/mood disorder

Using SDQ scales to distinguish between the community sample and clinical sample diagnosed with CD/ODD
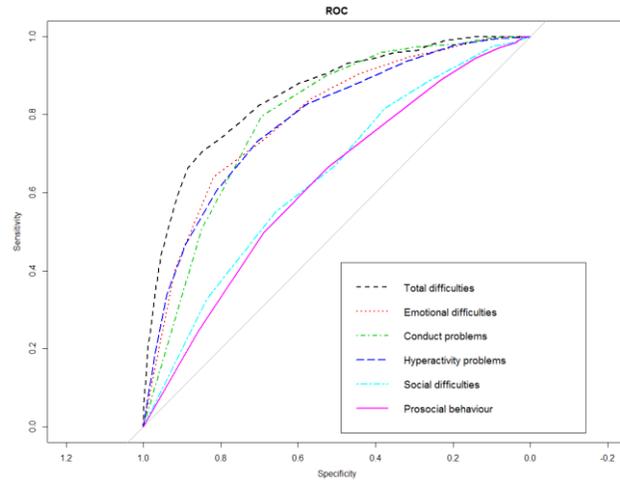
Fig. S21 female adolescents

Fig. S23 female adolescents

Fig. S25 female adolescents

Fig. S22 male adolescents
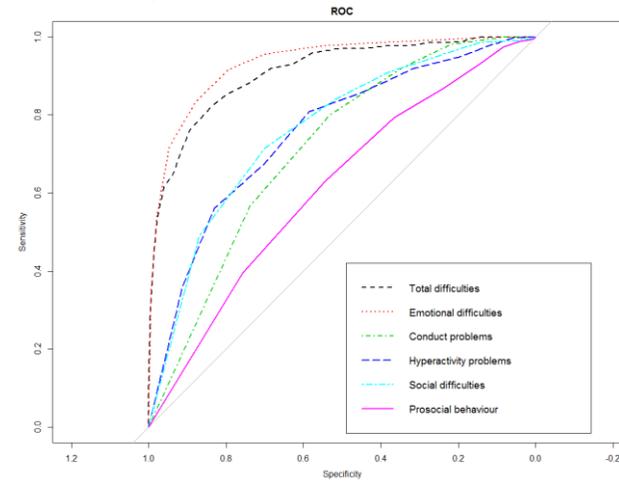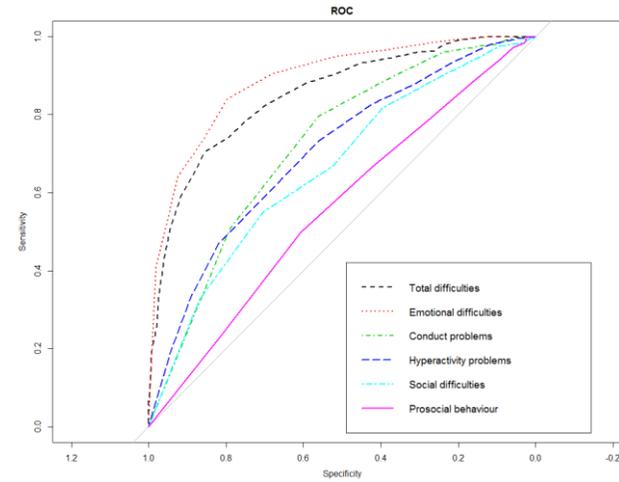
Fig. S24 male adolescents
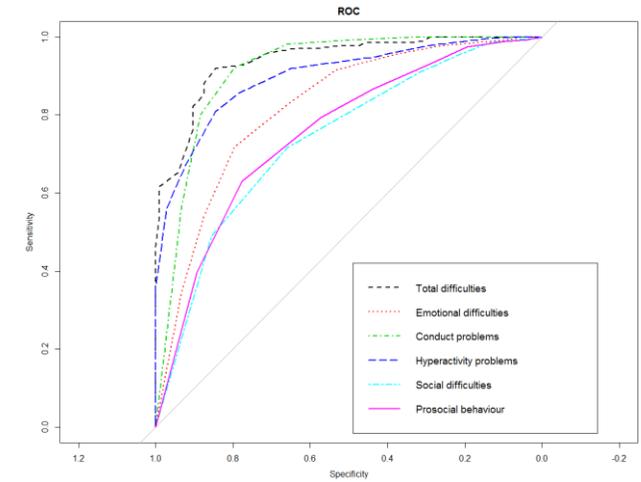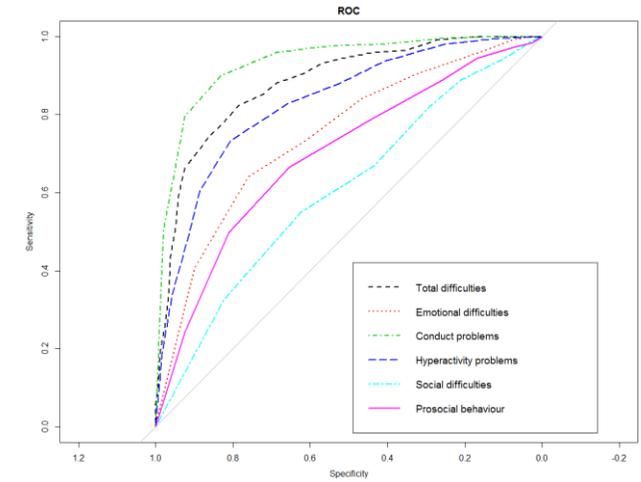
Fig. S26 male adolescents

Using SDQ scales to distinguish between the community sample and clinical sample diagnosed with ADHD
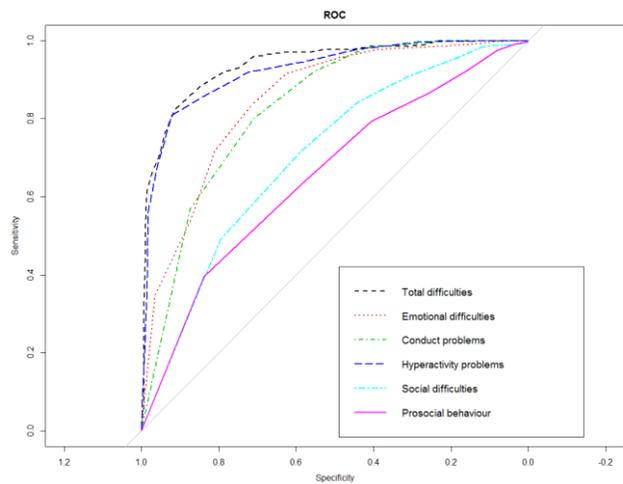

Fig. S27 female adolescents
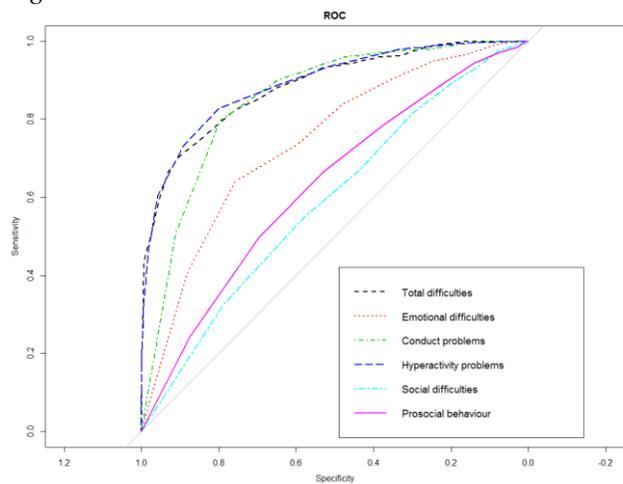
Using SDQ scales to distinguish between the community sample and clinical sample diagnosed with ASD


Fig. S29 female adolescents

Fig. S28 male adolescents
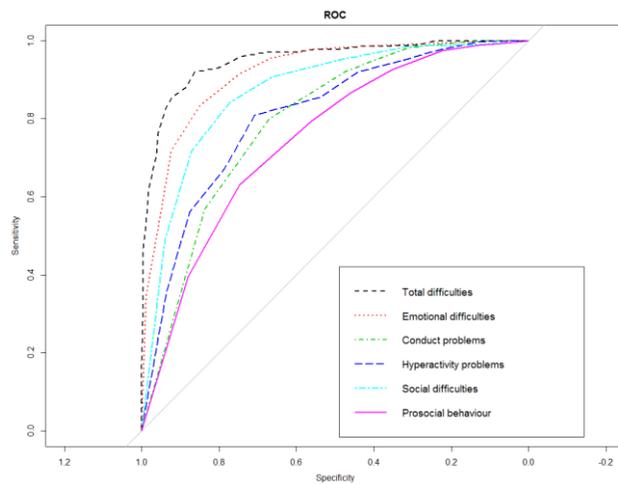
Fig. S30 male adolescents