ORIGINAL CONTRIBUTION

John A. Rønning
Bjørn Helge Handegaard
Andre Sourander
Willy-Tore Mørch

The Strengths and Difficulties Self-Report Questionnaire as a screening instrument in Norwegian community samples

■ **Abstract** This study reports on the application of the Norwegian self-report version of the Strengths and Difficulties Questionnaire (SDQ-S). The application of the SDQ-S was not motivated by a wish to reveal the prevalence of psychiatric disorders, but rather to set the stage for routine screening as part of schools' efforts to inform themselves about the life of adolescents at school. The survey included 4167 young people aged 11 to 16 years, attending 66 primary and secondary schools in Northern Norway. The respondents comprised 80.2% of the total population in these grades in the target area. Structural analysis of the instrument, including confirmatory factor analysis, internal consistency and intra- and cross-scale correlations revealed somewhat variable psychometric properties. Model modification suggested several ways of improving the structural psychometric properties of the SDQ-S. Norwegian cut-off points were similar to those found in other Scandinavian studies. About one third of the subjects reported at least minor perceived difficulties, while about 5% reported definite or severe difficulties. These difficulties were strongly associated with all symptom scales. Girls reported a significantly higher level of emotional problems and better prosocial functioning. Boys reported significantly higher scores on the externalising scales and on peer problems. The SDQ-S may be judged as an efficient and economical screening instrument for preventive research on large community samples. However, efforts should be made to improve its psychometric structure.

■ **Key words** strengths and difficulties questionnaire – self-report – screening – child psychiatry – prevention

J. A. Rønning (✉) · B. H. Handegaard · A. Sourander · W.-T. Mørch
Department of Child
and Adolescent Psychiatry
University of Tromsø
9016 Tromsø Norway
E-Mail: johnr@fagmed.uit.no

Introduction

Recent studies from the UK [7, 8], Germany [12], Sweden [19] and Finland [13] have generally reported acceptable psychometric properties for the "Strengths and Difficulties Questionnaire" (SDQ) in such areas as interrater agreement, internal consistency, and concurrent and discriminant validity. Three studies using various forms of exploratory factor analysis (EFA) [7, 14, 19] have confirmed the predicted five-factor structure (emotional, conduct, hyperactivity-inattention, peer problems, prosocial). However, EFA methods should be used primarily to discover factors rather than to test theories regarding the existence of factors. The types of analysis that are useful for testing theories regarding the existence of factors are referred to as confirmatory factor analysis (CFA), as distinct from EFA [15]. Furthermore, two of these studies [14, 19] uncovered somewhat low factor loadings on items on scales of conduct and peer problems. Further testing of its structural psychometric properties therefore seems warranted.

The SDQ Self-Report Questionnaire (SDQ-S) consists of only 25 core items plus a brief impact supplement that asks whether the respondent believes that he or she has a problem (perceived difficulties), and if so, inquires further about overall distress, social impairment (impact caseness), burden, and chronicity. The short-format SDQ-S is therefore an interesting alternative to longer instruments such as the Youth Self-Report (YSR; [1]) as a clinical screening instrument and in epidemiological and prevention research. This study has

three aims: 1) to report on the structural psychometric properties of the Norwegian version of the SDQ-S, 2) to report on prevalences and study changes in various problem domains across grade (6 through 10) and gender, and 3) to study associations between SDQ-S symptom scores and perceived difficulties and impact caseness.

Method

■ Sample

The study was carried out in 66 primary and secondary schools in three school psychology service districts in Northern Norway in spring 2001. The respondents made up 80.2 % of the total population of 5225 young people in sixth through tenth grade, aged 11 to 16 years; 2017 girls and 2150 boys.

■ Procedure

The project was approved by the Regional Ethical Committee and the Norwegian Data Inspectorate. The parents were contacted by the local school and given written and oral information about the project, and asked for their consent.

The questionnaires were administered in class by the class teacher, and students completed the questionnaires anonymously during a school lesson. The questionnaires were collected in a sealed envelope. Most of the non-responders were either absent from school or their parents or guardians had not given their consent to participation. No further descriptions of characteristics of non-participants were collected because of the promise of complete anonymity.

■ Measures

The SDQ-S includes 25 items divided into five scales of five items each: the hyperactivity scale, emotional symptoms scale, conduct problem scale, peer problem scale and prosocial scale. Each item is scored 0 for "not true", 1 for "somewhat true" and 2 for "certainly true", making it possible to sum the scores to generate a total difficulty score ranging from 0 to 40. The prosocial scale is a score of prosocial behavior and is thus not included in the total difficulties score. The cut-off point for clinical range is usually recommended to be roughly above the 90th percentile of total scores. The SDQ-S also includes an impact supplement of four point scales, in which the respondent is asked to what degree he/she perceives a problem, and if so, the duration or chronicity of the problem(s), its degree of impact and arenas of manifes-

tation, and the burden the problem(s) places on the respondent's social environment.

The SDQ-S was translated and backtranslated by Heiervang et al. (www.sdqinfo.com).

■ Statistical analysis

Multi-sample confirmatory factor analysis was performed using the LISREL 8.51 [11] in order to test and analyse the model proposed by Goodman [6, 7], to compute the correlations (ϕ) between the unobserved latent variables, and to test whether the factor loadings were significantly different for boys and girls. With large sample sizes, even small misspecifications in the model can lead to a significant chi-square. Other indices of fit should therefore be considered when evaluating the model fit for large samples. The comparative fit index (CFI), the Goodness-of-fit index (GFI), the standardized root mean square residual (SRMR), and the root mean square error of approximation (RMSEA) were used to evaluate model fit. The model evaluation was performed by inspection of modification indices and polychoric correlations (r^*). The modification index estimate decreases in chi-square when a particular change in the model is performed. Since SDQ-S variables are ordinal, the polychoric correlation matrix (computed by PRELIS 2.50) was used as a basis for the analysis. Parameters were estimated by the Weighted Least Squares (WLS) method, using the asymptotic covariance matrix (computed by PRELIS) in the weighting process, as recommended by Jöreskog and Sörbom [11].

SPSS 10 for Windows (release 10.0.5) was used to study basal statistics and internal consistency (Cronbach's alpha). When assessing the associations between observed scale scores, Pearson correlation was used. Univariate analysis of variance (ANOVA) was used to compare mean scores for boys and girls across grades on total difficulties and in each of the domains. Since the distribution of most of the dependent variables was fairly skewed, a logarithmic (natural log) transformation on these variables was employed to make the violation of the normality assumption less problematic.

Due to the number of comparisons and the sample size, we took $p < 0.001$ as the significance level, unless otherwise specified. To test the degree of difference between various population parameters, Cohen's [2] various types of effect sizes (ES) were chosen. Associations between gender, grade and symptom scores, and perceived difficulties and impact caseness were studied by means of univariate and multivariate logistic regression analysis. We expected that young people who scored high on the SDQ-S symptom scale would be more likely to perceive that they have definite or severe difficulties. In the univariate logistic regressions we evaluated each of the independent variables separately, including grade

J. A. Rønning et al.
SDQ self-report as a screening instrument

75

and gender, and whether these predicted the respondents' perceived difficulties. In the multivariate logistic regression analysis, the effects of each independent variable controlling for the others were studied. Odds ratios (OR) and 95% confidence intervals (95CI) were calculated.

Results

■ Psychometric properties

Factor analysis

Multi-sample confirmatory factor analysis was applied to test whether the factor loadings were significantly different for boys and girls (Table 1). Two models were compared: one in which the factor loadings were kept invariant between boys and girls, and one in which they were allowed to vary. The difference in chi-square be-

Table 1 Standardised factor loadings (weighted least squares estimates) and Cronbach's alpha of SDQ self-report for girls (G) and boys (B)

SDQ items	Prosocial		Emotion		Hyperactivity		Peer		Conduct	
	G	B	G	B	G	B	G	B	G	B
Factor loadings										
Prosocial behaviour (Cronbach's alpha: boys – 0.62, girls – 0.66)										
1. I try to be nice to other people. I care about their feelings	0.73	0.71								
4. I usually share with others (food, games, pens etc.)	0.52	0.48								
9. I am helpful if someone is hurt, upset or feeling ill	0.58	0.58								
17. I am kind to younger children	0.46	0.47								
20. I often volunteer to help others (parents, teachers, children)	0.51	0.52								
Emotional problems (Cronbach's alpha: boys – 0.64, girls – 0.67)										
3. I get a lot of headaches, stomach-aches or sickness			0.50	0.48						
8. I worry a lot			0.60	0.61						
13. I am often unhappy, down-hearted or tearful			0.79	0.76						
16. I am nervous in new situations. I easily lose confidence			0.58	0.57						
24. I have many fears. I am easily scared			0.62	0.62						
Hyperactivity (Cronbach's alpha: boys – 0.66, girls – 0.67)										
2. I am restless, I cannot stay still for long					0.71	0.71				
10. I am constantly fidgeting or squirming					0.67	0.66				
15. I am easily distracted. I find it difficult to concentrate					0.74	0.74				
21. I think before I do things					−0.46	−0.45				
25. I finish the work I am doing. My attention is good					−0.68	−0.70				
Peer problems (Cronbach's alpha: boys – 0.58, girls – 0.61)										
6. I am usually on my own. I generally play alone or keep to myself							0.59	0.56		
11. I have one good friend or more							−0.61	−0.65		
14. Other people my age generally like me							−0.56	−0.55		
19. Other children or young people pick on me or bully me							0.71	0.73		
23. I get on better with adults than with people my own age							0.49	0.47		
Conduct problems (Cronbach's alpha: boys – 0.54, girls – 0.44)										
5. I get very angry and often lose my temper									0.57	0.59
7. I usually do as I am told									−0.49	−0.52
12. I fight a lot. I can make other people do what I want									0.55	0.52
18. I am often accused of lying or cheating									0.61	0.65
22. I take things that are not mine from home, school or elsewhere									0.35	0.40

tween the two models was 32.68 with 25 degrees of freedom (p = 0.15), indicating that there was no difference in factor loadings between boys and girls.

Concerning the model fit, the GFI provides a measure of the relative amount of variance and covariance accounted for by the model, whereas the CFI provides a measure of fit of a particular model relative to another model, usually a null model. Values higher than 0.90 for the GFI and CFI are usually required to indicate a good fit. The SRMR represents the average of the absolute discrepancies between the observed and hypothesized matrices in correlational units. Values of 0.08 or lower have been suggested as necessary to consider a model fit to be acceptable. RMSEA provides a measure of model fit relative to the population covariance matrix when the complexity of the model is taken into account. Values lower than 0.05 for the RMSEA indicate a close fit, with values as high as 0.08 representing a reasonable fit [10].

As expected, the chi-square test was significant (Chisquare = 2368.8 (df = 265) p < 0.0001. The CFI, GFI, SRMR, and RMSEA values for the Goodman model were 0.82, 0.98, 0.11, and 0.47, indicating a somewhat variable and questionable fit.

Internal consistency

Cronbach alphas for the prosocial behavior, emotional problems, hyperactivity and peer problem subscales for boys and girls were in the range 0.58–0.67. The lowest Cronbach alphas were found on the conduct problem scale (0.44 for girls and 0.54 for boys).

Intra-scale correlations

On the conduct scale, very small correlations ($r* \leq 0.20$) were found, for both boys and girls, between item 7 ("usually do as I am told") and items 22 ("take things that are not my own from home, school and elsewhere"), 18 ("often accused of lying or cheating") and 12 ("I fight a lot. I can make other people do what I want"). For boys a very small correlation was also revealed between item 7 and item 5 ("get very angry and often lose my temper").

On the peer problem scale, very small correlations, for both boys and girls, were discovered between item 11 ("I have one good friend or more") and item 23 ("I get on better with adults than with people of my own age") and between item 11 and item 14 ("other people of my age generally like me"). For boys there was a very small correlation between item 11 and item 19 ("other children or young people pick on me or bully me").

On the hyperactivity scale very small correlations, for both boys and girls, were discovered between item 15 ("easily distracted. I find it difficult to concentrate") and item 21 ("I think before I do things"), and between item 10 ("constantly fidgeting or squirming") and item 21.

For boys a very small correlation was also found between item 15 and item 2 ("restless, I cannot stay still for long"). Often, very small intra-scale correlations were related to oppositely worded questions (positive vs. negative).

Cross-scale correlations

Cross-scale phi (φ) and Pearson (r) correlations for boys and girls combined showed the following: the strongest correlations were found between conduct problems and hyperactivity, φ = 0.83 (r = 0.49), and between conduct and peer problems, φ = 0.73 (r = 0.29). Between emotional problems and peer problems the association was φ = 0.77 (r = 0.40). Correlations between hyperactivity and emotional problems were φ = 0.59 (r = 0.29), and between conduct problems and emotional problems φ = 0.55 (r = 0.24). Conduct problems, peer problems and hyperactivity all correlated moderately well with prosocial behaviour, φ = –0.40 to –0.58 (r = –0.15 to –0.32). A moderate correlation was also found between peer problems and hyperactivity, φ = 0.46 (r = 0.17). Between prosocial behaviour and emotional problems there was a small to negligible correlation of φ = –0.19 (r = –0.09).

Further model analysis and modification

The Modification Index (MI) is used to suggest potential improvements to the model. The MI is employed to measure how much the chi-square is reduced when a particular change in model is suggested. The large sample size can make even a small MI significant. Therefore, only the largest MI is analysed in this text.

The modification indices that suggest an additional error covariance between items revealed a very large MI between item 2 (I am restless, I cannot stay still for long) and item 10 (I am constantly fidgeting and squirming) MI = 272. This indicates that these two items create a sub-dimension within the Hyperactivity factor (restless, fidgeting). In the Norwegian language there is a close semantic similarity between these two items that may explain the very high correlation of $r* = 0.70$.

Some MI point to possible relationships between already defined latent factors and indicators of other latent factors. To improve model fit, one can adjust the model by allowing more paths using MI, and adding one path at a time. Of course, only theoretically relevant items should belong to the same latent construct.

The MI suggests adding a path from item 21 (I think before I do things; Hyperactivity) to Prosocial Behaviour (MI = 107). This may be an effect of the positive wording of item 21. Re-estimating the model with item 21 linked to both Hyperactivity and Prosocial Behaviour, suggests that a path could be added from item 13 (I am often unhappy, downhearted or tearful; Emotional Problems) to Peer Problems; MI = 82. Although item 13

J. A. Rønning et al.
SDQ self-report as a screening instrument

77

is not a direct indicator of Peer Problems, this item is often correlated with items belonging to this construct. $r*$ between item 13 and item 19 (Other children or young people pick on me or bully me; Peer Problems) for the whole sample is 0.54.

The next step would link item 15 (I am easily distracted, I find it difficult to concentrate; Hyperactivity) to Emotional Problems; MI = 80. Item 15 is moderately correlated with all five items within the Emotional Problems scale ($r*$ between 0.29 and 0.42).

The fourth step would link item 18 (I am often accused of lying or cheating; Conduct Problems) to Peer Problems; MI = 89.

The fifth step would link item 5 (I get very angry, and often lose my temper; Conduct Problems) to Emotional Problems; MI = 128.

The sixth step would link item 14 (Other people my age generally like me; Peer Problems) to Prosocial Behaviour; MI = 74.

More modifications could be made, but the number of changes should be limited.

When the effects of modifications with CFA were tested, the following results emerged: Chi-square = 1487.7 (df = 258), CFI = 0.89, GFI = 0.99, SRMR = 0.075, RMSEA = 0.036, indicating an improved fit.

■ Prevalences of symptoms, mean scores and cut-off points

Item scores

In general, these young people reported more strengths than difficulties. Table 2 reveals a high level of problems (> 10 % prevalence) among girls: in the emotional domain, pain symptoms, worries, and social nervousness (items 3, 8, 16), and in the hyperactivity domain for four out of five items, restlessness, fidgeting or squirming,

Table 2 The wording of items and endorsement rates (%) of response categories in the *severe area of the SDQ self-report for girls (G) and boys (B) in the Norwegian sample

	Severe		
	G	B	ES[1]
Prosocial behaviour			
1. I try to be nice to other people. I care about their feelings	0.6	3.4	0.22
4. I usually share with others (food, games, pens etc.)	2.6	8.3	0.26
9. I am helpful if someone is hurt, upset or feeling ill	1.5	7.9	0.32
17. I am kind to younger children	3.2	7.9	0.21
20. I often volunteer to help others (parents, teachers, children)	7.5	15.2	0.25
Emotional symptoms			
3. I get a lot of headaches, stomach-aches or sickness	14.9	9.2	0.18
8. I worry a lot	13.6	11.2	0.07
13. I am often unhappy, down-hearted or tearful	7.9	4.5	0.14
16. I am nervous in new situations. I easily lose confidence	13.7	11.6	0.06
24. I have many fears. I am easily scared	8.6	4.7	0.16
Hyperactivity			
2. I am restless, I cannot stay still for long	12.7	15.5	0.08
10. I am constantly fidgeting or squirming	12.8	17.5	0.13
15. I am easily distracted. I find it difficult to concentrate	12.8	15.6	0.08
21. I think before I do things	6.6	10.0	0.12
25. I finish the work I am doing. My attention is good	11.1	14.0	0.09
Peer problems			
6. I am usually on my own. I generally play alone or keep to myself	9.1	10.4	0.04
11. I have one good friend or more	2.1	2.4	0.02
14. Other people my age generally like me	5.4	9.1	0.14
19. Other children or young people pick on me or bully me	3.8	6.3	0.12
23. I get on better with adults than with people my own age	7.0	10.0	0.11
Conduct problems			
5. I get very angry and often lose my temper	10.1	13.6	0.11
7. I usually do as I am told	3.7	5.5	0.09
12. I fight a lot. I can make other people do what I want	1.5	4.2	0.17
18. I am often accused of lying or cheating	4.9	10.9	0.23
22. I take things that are not mine from home, school or elsewhere	3.3	3.9	0.03

* Severe is defined as "Not True" for positively directed questions, and "Certainly True" for negatively directed questions
[1] *ES* effect size

distraction and problems of attention (items 2, 10, 15 and 25); and in the conduct domain, lack of inhibitory control (item 5). Similarly, boys reported a high level of problems in all hyperactivity items, problems with impulsiveness in addition to the same problems as the girls (items 2, 10, 15, 21 and 25).

Boys also reported a high level of problems (> 10 % prevalence) in the conduct domain as regards lying and cheating and inhibitory control (items 5 and 18); in the prosocial domain with helping others (item 20) and in the peer problem domain regarding social withdrawal and with similarly aged peers (items 6 and 23). Boys also reported high emotional symptoms in areas such as worries and social nervousness (items 8 and 16).

Applying Cohen's [2] criteria for effect size between boys and girls, all ES fell in the low area (< 0.29) except for item 9, which fell in the moderate area, ES = 0.32.

Mean scores

Except for total scores, across all domains the effect of gender was significant at $p < 0.001$. As shown in Table 3, girls reported a higher level of emotional symptoms, and a better level of prosocial functioning. Boys reported a higher level of problems concerning hyperactivity, conduct and peer problems. The effect of grade was significant on all domains except for emotional symptoms. However, differences across grades and between gender were less than one point in all domains. The gender-grade interaction was significant in prosocial behaviour ($p < 0.01$), and in emotional symptoms ($p < 0.01$), indicating that girls reported growing emotional problems, while boys reported a slightly lower level of prosocial functioning at higher grade levels.

Cut-off points

As in other studies that have reported cut-off-points of the SDQ-S (e. g. 15), it was difficult to decide exactly on the 80th or 90th percentile for problem scales or the 20th or 10th percentile on the prosocial scale (Table 4). Generally speaking, more boys than girls scored above the 90th percentile on all problem scales except on the emotional scale.

■ Impact measures

Prevalence of impact measures

Altogether 1235 (36 %) subjects reported either minor difficulties (31.5 %), definite difficulties (3.6 %) or severe difficulties (0.9 %). Of those reporting at least minor difficulties 26.4 % reported that these had lasted less than

Table 3 Total and sub-scale scores for the SDQ-S (mean values and (SD)) by gender and grade

Grade	6		7		8		9		10	
Gender N	Girls 514	Boys 563	Girls 498	Boys 518	Girls 365	Boys 385	Girls 334	Boys 697	Girls 290	Boys 301
Total problems	8.90 (5.13)	9.99 (5.96)	8.82 (5.41)	9.59 (5.32)	9.20 (4.80)	9.70 (5.57)	10.05 (5.39)	9.91 (5.27)	10.59 (5.51)	10.20 (5.35)
Hyperactivity	3.08 (2.04)	3.54 (2.11)	3.13 (2.03)	3.68 (2.19)	3.47 (2.08)	3.63 (2.23)	3.81 (2.25)	3.99 (2.39)	4.10 (3.43)	4.00 (2.30)
Emotional symptoms	2.72 (2.21)	2.22 (2.14)	2.58 (2.27)	1.95 (1.92)	2.68 (1.97)	2.02 (1.83)	3.06 (2.35)	1.86 (1.90)	3.01 (2.35)	1.96 (1.88)
Conduct problems	1.39 (1.31)	1.97 (1.74)	1.46 (1.43)	1.99 (1.59)	1.56 (1.44)	2.13 (1.84)	1.62 (1.42)	2.33 (1.82)	1.84 (1.65)	2.27 (1.88)
Peer problems	1.70 (1.64)	2.26 (1.90)	1.66 (1.84)	1.99 (1.81)	1.50 (1.62)	1.92 (1.91)	1.52 (1.60)	1.75 (1.78)	1.63 (1.56)	1.96 (1.84)
Prosocial behaviour	8.15 (1.65)	7.03 (2.01)	7.90 (1.62)	6.79 (1.88)	7.72 (1.66)	6.55 (1.93)	7.77 (1.67)	6.11 (2.21)	7.70 (1.69)	6.21 (1.97)

Table 4 Norwegian cut-off points of the SDQ self-report total difficulties score and subscores*, and percentage of girls (G) and boys (B) (11–16 year olds) scoring in the borderline (> 80th < 90th percentile) and abnormal (> 90th percentile)

	Borderline range (80th–90th percentile)				Clinical range (above 90th percentile)			
	Cut-off	Total %	G%	B%	Cut-off	Total %	G%	B%
Hyperactivity	6	8.6	8.2	9.1	7	10.0	8.8	11.1
Emotional symptoms	5	7.1	9.1	5.1	6	9.4	12.5	6.5
Conduct problems	4	7.8	5.9	9.6	5	7.0	4.1	9.8
Peer problems	4	7.0	5.9	8.1	5	8.8	7.2	10.4
Total problems	15	8.9	9.0	8.8	18	9.0	7.6	10.4
Prosocial behaviour	5	11.5	7.3	15.6	4	8.3	3.0	13.2

* On the prosocial subscale, the percentile range for borderline through clinical is inverted compared to the other scales

J. A. Rønning et al.
SDQ self-report as a screening instrument

79

one month, 21.1% one to five months, 10.3% 6–12 months, and 42.2% more than one year. Furthermore, applying only the ranges "Quite a lot" and "A great deal" of the scale, 28.4% of the sample reported that they were distressed by the difficulties, while 21.6% reported that the difficulties interfered with their home life, 18.8% with friendship, 29.6% with classroom learning and 14% with leisure activities. 13.6% reported that their difficulties put "Quite a lot" to "A great deal" of burden on their general social milieu.

Perceived difficulties, symptom scales and impact caseness

The associations of gender, grade and dichotomised problem domains with perceived difficulties (Low = no and minor problems, n = 3247, and High = at least definite or severe problems, n = 152) were studied using univariate logistic regression analysis. No significant association was found on gender and age. As shown in Table 4, SDQ-S total scores, emotional problems, conduct

problems, hyperactivity, peer problems, and prosocial behaviour problems were all associated with high levels of perceived difficulties. Because of the hierarchical relationship between total scores and subscores, only subscores were included in the further multiple regression analysis, which indicated that a high level of emotional problems had the strongest association with perceived problems (OR 6.0, 95% CI 4.1–8.8). Hyperactivity (OR 2.0, 95% CI 1.3–3.0), conduct (OR 2.3, 95% CI 1.4–3.6), and peer problems (OR 2.5, 95% CI 1.6–3.7) were also independently associated with perceived difficulties. Of those with definite or severe perceived difficulties, 47.4% were above the 90th percentile in the SDQ-S total scores and 65.8% above the 80th percentile.

The associations of gender, grade and problem domains on impact caseness, applying Goodman's [6] recommendations about adding the scores on the distress and social incapacity items using a "0012" scale and an impact cut-off score of 2, were studied by means of univariate logistic regression analysis. A total of 399 subjects scored two points or more on impact caseness. No

Table 5 Associations between sex, grade, SDQ-S symptom domains, and perceived difficulties and impact caseness. Odds ratio with asymptotic 95% confidence interval. The p-value is based on the (univariate) test of the hypothesis that the population odds ratio is different from 1. N = 3435

Variable	n	% perc. diff.	% impact caseness	Perceived diff. Odds ratio (95% CI)	Impact caseness Odds ratio (95% CI)
Gender					
girls	1662	4.3	13.1		
boys	1755	4.7	10.3	1.1 (0.8–1.5)	0.8 (0.6–0.95)
Age					
11	881	12.7			
12	828	3.3	8.6		
13	598	3.7	10.4		
14	594	4.9	12.6		
15	534	6.4	14.8		
				1.1 (0.99–1.2)	1.07 (0.99–1.15)
Total problems					
< 90%	3093	2.6	8.2		
> 90%	325	22.2	43.7	10.7* (7.6–15.1)	8.6* (6.7–11.1)
Hyperactivity					
< 90%	3060	3.5	9.5		
> 90%	371	12.7	28.6	4.0* (2.8–5.8)	3.8* (2.9–4.9)
Emotional problems					
< 90%	3084	2.8	8.6		
> 90%	338	19.5	39.1	8.4* (5.9–11.8)	6.8* (5.3–8.8)
Conduct problems					
< 90%	3171	3.6	9.9		
> 90%	251	15.5	33.9	4.9* (3.3–7.3)	4.7* (3.5–6.2)
Peer problems					
< 90%	3099	3.3	9.2		
> 90%	321	15.3	34.3	5.2* (3.6–7.5)	5.1* (4–6.7)
Prosocial behaviour					
< 10%	3124	4.1	11.5		
> 10%	293	8.2	13.0	2.1* (1.3–3.3)	1.2 (0.8–1.6)

* p < 0.001

significant association was found with age, but a small but significant sex difference was found, more girls reporting impact caseness than boys. As Table 5 shows, SDQ-S total scores, emotional problems, conduct problems, hyperactivity, and peer problems were all associated with impact caseness. No significant association was found between prosocial behavior and impact caseness.

Multiple regression analysis including only subscores indicated that the domains that demonstrated the strongest significant associations with impact caseness were emotional problems (OR 4.2, 95 % CI 3.2–5.6), conduct problems (OR 2.8, 95 % CI 2.0–3.9), peer problems (OR 3.2, 95 % CI 2.3–4.3) and hyperactivity (OR 2.4, 95 % CI 1.8–3.3) all p's < 0.00005. A small but significant association was also observed with gender (OR 0.73, 95 % CI 0.6–0.9) p = 0.01. Of those reporting at least two points on the impact score, 35.8 % were above the 90th percentile in the SDQ-S total scores and 58.4 % above the 80th percentile.

Discussion

Psychometric properties

The overall picture that emerged when the SDQ-S was subjected to this range of psychometric tests questioned its psychometric structural properties. The CFA generally confirmed that the five-factor Goodman model was similar for boys and girls. However, the goodness-of-fit indices of the original model indicated a range of fits, with only the GFI and RMSEA showing acceptable fit.

Inspecting the Cronbach alphas and intra-scale correlations, low alphas were revealed in the domains of peer problems and conduct problems. Similar problems (alpha < 0.6) have been reported on the peer problem scale in other studies [7, 14], and for the conduct problem scale, alpha < 0.7 [7, 14, 16]. Low alphas indicate that the items included were measuring somewhat different properties. Common to these domains in this study were very small intra-scale correlations between oppositely posed questions. This suggests a potential problem with the transformation of some of the positively worded items to negative scales. It could be that the respondents were not interpreting these questions in the way intended by the designer of the questionnaire. It might well be that changing the formulation of these questions from positive to negative would increase both the Cronbach alphas and the model fit.

Model analysis and modification

When the Modification Index analysis was used to modify the Goodman model, some improvements in the fit indices were obtained, although these were not enough to provide a perfect fit on all indices. Nor is it desirable to have many items loading on two or three latent variables. Thus, the meaning of the constructs in the modified model might be different from those in the original model. However, it could be argued theoretically that an item ought to belong to more than one latent construct. We find that the residuals (differences between observed and estimated correlations) are smaller in the modified model (smaller SRMR). Yet more adjustments would have to be made in order to satisfy the criteria for a perfect model. The polychoric correlations between many items not belonging to the same construct were moderate to high. This is not surprising, since many children could show more than one symptom at the same time. However, it might be worth reformulating some items (at least in the Norwegian version) to make more distinct dimensions. To do so, it would be important to avoid semantic similarities, eliminate positive and negative wording within the same construct, and avoid two sentences within the same item.

Symptoms and cut-offs

The cut-off points and mean scores on the symptoms domains were generally somewhat lower than those reported from the UK [5], although they were on a rather similar level to those reported from Finland [13, 14]. For example, the 90th percentile cut-off point for total scores was 18 points in the present and Finnish samples [14], whereas in the UK sample it was 20 points. Specific norms adjusted for age and gender should be employed in different cultures and in different samples, as previously recommended [3, 6, 8].

When the effects of age and sex were studied in the different individual problem domains and for prosocial behaviour, the results were as expected. Girls reported higher levels of emotional symptoms and better prosocial behaviour. This difference increased in the higher grades. Boys reported higher level of externalising problems across grades. The SDQ cut-off points of different scales covered a significantly different percentage of boys and girls. The percentile cut-off points may be used, for example if the researcher wishes to know whether a sample of girls with high SDQ scores is comparable to age- and sex-matched controls. However, Goodman originally suggested that the cut-off points of the SDQ could act as useful "warning signals" to the clinician that an adolescent might have a disorder [5, 8]. Thus, cut-off points may act as markers for preventive or clinical action.

The SDQ-S includes a peer problem domain that is found exclusively in this instrument. The importance of peer friendship for children's sense of wellbeing and adjustment, and lack of such friendship for maladjustment

J. A. Rønning et al.
SDQ self-report as a screening instrument

81

and psychopathology, is well documented [4]. In the present study boys reported more problems in peer relationships and less prosocial behaviour across grades. This result is in accordance with previous reports (e. g. [14]). However, the number of subjects in this study was large, and that should be considered when interpreting the significances. Although sex differences were statistically highly significant, the actual differences were less than one point, which is less than 10 percent of the range of some individual subscales.

Perceived difficulties and impact caseness

About one-third of these school-age children reported at least minor difficulties and about 5 % reported perceived definite or severe difficulties. The difficulties tended to have an impact on several areas of the child's social environment, such as home, peer relationships, school and leisure activities. Furthermore, about 12 % of the subjects had an impact caseness score of 2 or more. Similar results were reported by the Goodman [6] community sample from the UK. Traditionally, questionnaires have predicted psychiatric caseness on the basis of symptom scores. However, it has been reported that when the extended SDQ is employed, clinical status is better predicted by impact than by symptoms [6]. Interestingly, Goodman [6] reported that a psychiatric disorder can be predicted rather well from a single question on perceived difficulties, and only a little less accurately than caseness predicted by the combination of symptoms and impact.

Both the perceived definite or severe difficulties and the impact caseness were independently associated with all symptom scales except prosocial behavior. Emotional problems had the strongest association with perceived difficulties and impact caseness. However, only about half of those who reported definite or severe difficulties were above the 90th percentile cut-off point in the total scores. One explanation for this is that adolescents may have severe and disabling symptoms in areas not covered at all or not very well in the SDQ (e. g. dieting problems, panic attacks, fears, psychotic symptoms) and therefore have low SDQ symptom scores.

Limitations and clinical implications

This study might be criticised for not reporting test-retest reliability. However, as Nunnally [15] has pointed out, if a coefficient alpha is low for a test, a relatively high correlation between retests should not be taken as an indication of high reliability. What is important is that the test correlates with itself. On these grounds he recommends that the retest method should generally not be used to estimate reliability.

Another source of weakness of this study is the lack of information about the characteristics of the students who did not participate. It is possible that some of these were students who were experiencing problems (e. g. [16]). However, the participation rate was high. There are reasons to suppose that the participation rate would have been lower without the promise of complete anonymity [15, 16].

A greater limitation of this study is the lack of cross-informant comparisons, since only the SDQ self-report questionnaire (SDQ-S) was employed. Subjective appraisals (self-reports) are more prone to distortions arising from intrapersonal factors (e. g. social desirability, over- or underestimation of the frequency or severity of events) than external observers' assessments. Goodman et al. [9] demonstrated in their large community sample that the full combination of the three-informant SDQ (parent, teacher, and self) questionnaires yielded the highest sensitivity in predicting psychiatric disorders. In their study, self-report was found to be the least useful single strategy, being less sensitive than parent report for all disorders, and less sensitive than teacher report for all disorders except depression. Obviously, when the SDQ is employed as a screening instrument for "true" clinical disorders at least two informants should be used. Additionally, the use of clinical interviews would have given us the possibility of addressing the sensitivity and specificity of the instrument.

As Goodman et al. [9] have stressed, when performing studies such as this, we must take into account certain ethical considerations, balancing a wish to investigate the clinical predictive ability of the instrument against the availability of effective treatment services once a diagnostic label has been decided on. In our case the SDQ-S as a screening device was employed as one measure for determining the overall well-being of a large group of young people in 66 primary and secondary schools. The application of the SDQ-S was not motivated by a wish to reveal the prevalence of psychiatric disorders, but rather to set the stage for routine screening as part of schools' efforts to inform themselves about adolescent life at school. In Norway, schools use several programs with the primary aim of improving the psychosocial conditions of young people. Screening instruments such as the SDQ-S are an efficient and economical means of gathering important information. The present study has also shown that the psychometric properties of the instrument in the Norwegian context are questionable, and a few suggestions for possible improvements are made. These would entail avoiding semantic similarities, positive and negative wording within the same construct, and two sentences within the same item. It might also be worth linking some of the items to more than one latent construct.

References

1. Achenbach TM (1991) Manual for the Youth Self-Report. Burlington, VT: University of Vermont, Department of Psychiatry
2. Cohen J (1988) Statistical power analysis for the behavioral sciences (Second edition). Hillsdale, New Jersey. Lawrence Erlbaum Associates
3. Crinjen AAM, Achenbach TM, Verhulst FC (1997) Comparisons of problems reported by parents of children in 12 cultures: Total problems, externalising, and internalising. J Am Acad Child Adolesc Psychiatry 36:1269–1277
4. Deater-Deckard K (2001) Annotation: Recent research examining the role of peer relationships in the development of psychopathology. J Child Psychol Psychiatry 42:565–579
5. Goodman R (1997) The Strengths and Difficulties Questionnaire: a research note. J Child Psychol Psychiatry 38: 581–586
6. Goodman R (1999) The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. J Child Psychol Psychiatry 40: 791–799
7. Goodman R (2001) Psychometric properties of the Strengths and Difficulties Questionnaire. J Am Acad Child and Adolesc Psychiatry 40:1337–1345
8. Goodman R, Meltzer H, Bailey V (1998) The strengths and difficulties questionnaire: A pilot study on the validity of the self-report version. Europ Child and Adolesc Psychiatry 7:125–130
9. Goodman R, Renfrew D, Mullick M (2000) Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. Europ Child Adolesc Psychiatry 9:129–134
10. Jöreskog K, Sörbom D (1993) LISREL 8: Structural equation modelling with the SIMPLIS command language. Chicago: SSI Scientific Software International
11. Jöreskog K, Sörbom D (1996) LISREL 8: User's reference guide. Chicago: SSI Scientific Software International
12. Klasen H, Woerner W, Wolke D, Meyer R, Overmeyer S, Kasnitz W, Rothenberger A, Goodman R (2000) Comparing the German versions of the Strengths and Difficulties Questionnaire (SDQ-Deu) and the Child Behavior Checklist. Europ Child Adolesc Psychiatry 9:271–276
13. Koskelainen M, Sourander A, Kaljonen A (2000) The Strengths and Difficulties Questionnaire among Finnish school-aged children and adolescents. Europ Child Adolesc Psychiatry 9:277–284
14. Koskelainen M, Sourander A, Vauras M (2001) Self-reported strengths and difficulties in a community sample of Finnish adolescents. Europ Child Adolesc Psychiatry 10:180–185
15. Nunnally JC (1978) Psychometric Theory. McGraw-Hill Book Company, New York
16. Nøvik TS (1995) Are postal surveys in child psychiatry feasible? Results of an epidemiological study. Nordic J Psychiatry 49:337–342
17. Ong AD, Weiss DJ (2000) The impact of anonymity of responses to sensitive questions. J Applied Soc Psychol 30: 1691–1708
18. Slee PT (1994) Situational and interpersonal correlates of anxiety associated with peer victimization. Child Psychiatry Human Develop 25:97–107
19. Smedje H, Broman JE, Hetta J, von-Knorring AL (1999) Psychometric properties of a Swedish version of the "Strengths and Difficulties Questionnaire." Europ Child Adolesc Psychiatry 8:63–70