A population-based psychometric validation study of the Strengths and Difficulties Questionnaire – Hebrew version

Ivonne Mansbach-Kleinfeld[1], Alan Apter[2]*, Ilana Farbstein[3], Stephen Z. Levine[4] and Alexander M. Ponizovsky[1]

[1] Mental Health Services, Ministry of Health, Jerusalem, Israel
[2] Department of Child and Adolescent Psychiatry, Schneider Medical Center for Children in Israel, Petach Tikvah, Israel
[3] Department of Child and Adolescent Psychiatry, Ziv Hospital, Zfat, Israel
[4] Department of Criminology, Bar Ilan University, Ramat Gan, Israel

This study presents the psychometric properties of the Strengths and Difficulties Questionnaire – Hebrew version (SDQ-H), used in the Israel Survey on Mental Health among Adolescents (ISMEHA). The SDQ-H was administered to a representative sample of 611 adolescents and their mothers. Structural validity was evaluated by exploratory and confirmatory factor analysis and the Development and Well-Being Assessment (DAWBA) inventory was used as "gold standard" to test convergent and discriminant validity. Internal consistency and normative scores were established. Agreement was found with the original factor structure, except for the Peer problem scale. Concurrent and discriminant validity varied from fair to very good for most scales. Total Difficulties scores showed better discriminant validity for the adolescents' than the mothers' report for internalizing disorders, and the opposite for externalizing disorders. Internal consistency for the Total Difficulties was 0.77 and for the Hyperactivity scale it was 0.73. It was lower for the other scales, particularly for the Peer problems scale. The findings suggest reasonable psychometric properties of the SDQ-H. Comparisons with other translated SDQ versions are presented.

Keywords: SDQ-Hebrew version, validity, reliability, normative scores

INTRODUCTION

The Strengths and Difficulties Questionnaire (SDQ) is a multi-informant screening measure designed to detect mental health problems in children and adolescents (Goodman, 1999). It has parent- and teacher-report versions for 4- to 16-year-olds, and a self-report version for 11- to 17-year-olds. The tool is increasingly being used in community and clinical settings and in cross-cultural research thanks to its brevity, accessibility, and availability in the public domain (http://www.sdqinfo.com).

Studies carried out in diverse settings support the SDQ's validity and reliability (Goodman, 2001; Alyahri and Goodman, 2006; Vostanis, 2006) and its predictive power across cultures, languages, and socio-economic backgrounds (Goodman et al., 2000a). Most report Cronbach's alpha coefficients above 0.70 for the Total Difficulties score (van Widenfelt et al., 2003), but low coefficients for the Conduct (Koskelainen et al., 2001), and Peer problems scales (Goodman, 2001).

Structural validity varies with some studies supporting the original five-factor structure (Hawes and Dadds, 2004; Woerner et al., 2004a), and others showing different patterns (Thabet et al., 2000; Koskelainen et al., 2001; Muris et al., 2003). Concurrent validity has been generally supported (Goodman and Scott, 1999; Muris et al., 2003) and SDQ scales have demonstrated fair to good discrimination between groups with and without Attention Deficit Hyperactivity Disorder (Yasong et al., 2008).

Regarding SDQ patterns and norms, it has been shown that populations that have a higher prevalence of mental disorders have higher mean scores in the SDQ scales (Goodman and Goodman, 2009). The prevalence of any disorder among Israeli adolescents aged 14–17 was reported as 11.7% (Farbstein et al., 2010), comparable to prevalence rates of mental disorders conducted worldwide, which have reported rates varying between 7 and 16.4% (Costello et al., 1997; Meltzer et al., 2003; Canino et al., 2004; Heiervang et al., 2007). Thus, we do not expect much variation in SDQ mean scores when comparing Israeli youth to that of the aforementioned countries. Cross-national differences should be assessed with the awareness that there may be reporting biases. For instance, a comparison of Norwegian and British parents and teachers found systematical under-reporting of emotional symptoms on the SDQ by Norwegian parents (Heiervang et al., 2008).

This paper presents the psychometric properties of the SDQ-Hebrew (SDQ-H) versions for parents and adolescents. It includes normative scores, internal consistency, and construct, concurrent and discriminant validity. Data were obtained from the Hebrew respondents of the Israel Survey of Mental Health among Adolescents (ISMEHA), conducted on a representative sample of Israeli adolescents and their mothers in 2004–2005 (Mansbach-Kleinfeld et al., 2010).

MATERIALS AND METHODS

THE SURVEY POPULATION

The ISMEHA collected data from a nationwide, cross-sectional representative sample of 14- to 17-year-old Israeli adolescents living in the community, in urban settings with more than 2,000 inhabitants and meeting the status of legal residents according to the National Population Register (NPR). Initially 13- to 15-year-olds were sampled

so that by time of interview they would be between 14 and 16 years of age. However, due to fieldwork constraints, not all the data was collected within schedule and therefore the age of the group shifted upward. Not included in the survey population were, adolescents residing in small rural settings, such as Kibbutzim or other collective settings (comprising 7.3% of the population in this age group), unrecognized Bedouin villages (1.7% in this age group); Palestinian residents of East Jerusalem (2.8% in this age group), and; Jewish ultra-orthodox (Haredi) adolescents (17.8% in this age group). The Haredi adolescents were excluded before data collection began due to low response rates achieved in the pilot stage (4%), despite active efforts made to adjust the survey's methods to the specific requests of ultra-orthodox leaders. The ultra-orthodox Jewish children and adolescents attend schools where girls and boys learn separately and live in more of less segregated neighborhoods in several cities. They are not allowed to watch TV or films, to read secular newspapers or use the web and thus parents and teachers have strong control over the information to which they are exposed. Therefore, the content of our survey was not deemed appropriate for this population by the religious leaders who were approached. The 245 Jewish ultra-orthodox adolescents excluded from the study were identified by the type of school they were attending and their home address, as they attend a separate educational system and live in specific neighborhoods. Children of migrant workers, a very small percentage of the population in this age group, were not included in the study, as most do not meet the status of legal residents in Israel.

THE SAMPLING FRAME
The sample was based on the NPR. The file, updated to August 2002, included the names of all residents, born in Israel or abroad, between July 1, 1987 and June 30, 1990; and demographic data such as home address, school of attendance, country of birth, and year of immigration, if relevant. New immigrants arriving in the country after August 2002 were not included in the sample. Deceased adolescents were removed from the sampling frame. The number of adolescents in this age group in the sampling frame (including Jewish ultra-orthodox adolescents) was 317,604.

SAMPLE SIZE AND SAMPLING PROBABILITY
Based on epidemiological studies that reported prevalence of mental disorders among adolescents of between 12 and 20% at the time the investigation was planned (Achenbach and Howell, 1993; Verhulst et al., 1997; Surgeon General, 1999; Canino et al., 2004), a sample size of 1,000 adolescents (without Jewish ultra-orthodox adolescents) was calculated to enable the identification of "any mental disorder" with adequate statistical power. Given an expected response rate of 2/3, an initial sample size of about 1,500 adolescents was chosen. The average sampling fraction was 1/212. However, calculations were made according to a sampling fraction of 1/190 and therefore the final sample was not 1500 but 1670, a little larger than originally planned.

SAMPLING METHOD
To increase the cost-efficiency of the study, we sampled localities with a minimum of 30 adolescents in this age group in the sample. All the urban settlements were distributed into sampling strata as follows:

1. Cities were ordered according to size of the adolescent population 13–15 years of age.
2. The three largest Israeli cities (Jerusalem, Tel Aviv, and Haifa) plus another eight large cities were included in the sample with certainty. The adolescents in those cities were chosen through systematic random sampling and they were sampled in one-step.
3. In all the other urban settlements the sampling was made in two-steps. The cities were distributed into strata according to the two main types of locality (whether they were Jewish/mixed or mainly Arab-populated cities) and by six geographical regions (Jerusalem District, Tel Aviv District, Haifa District, Northern District, Central District, and Southern District) and ordered within each stratum according to size. Size referred specifically to the estimated number of 13- to 15-year-olds in the sampling frame in each city.
4. The urban settlements in the sample were chosen through systematic random sampling with a probability proportional to size so that the final sample in each stratum represented all the adolescents in that stratum.
5. All adolescents within each of the sampled localities were ordered according to age, gender, and geographic distribution within the city, in an attempt to represent all socio-economic groups. The adolescents in that city were sampled by a systematic random sampling method. The sampling probability within each city was calculated so that the final sampling fraction would be the same for the total sample.

INSTRUMENTS
The Strengths and Difficulties Questionnaire
The original SDQ was translated into Hebrew, considering cultural aspects, colloquialism, and idioms to remove biases specific to the original tool. Bilingual researchers conducted the back-translation, which was compared to the original questionnaire (Goodman, 1999) and revised by the SDQ's author. Following a pilot study on 15 subjects a final decision on the exact wording was made.

The SDQ is a screening diagnostic instrument designed for evaluating social, emotional, and behavioral functioning in children and adolescents ages 4–17 years. It includes five subscales: four of them refer to difficulties and one to the adolescents' strengths. Its 25 items cover four clinical domains, namely, hyperactivity-inattention, emotional symptoms, peer-relationship problems, and conduct problems, and one distinct prosocial behavior domain. Each item/statement is rated on a 3-point scale as 0 (not true), 1 (somewhat true), or 2 (certainly true). The questionnaire is multi-informant, i.e., can be administered to adults (parents, caretakers, and teachers), and also includes a self-report version for adolescents aged 11–17. In addition to the clinical domains, the SDQ is supplemented with an impact module that asks the respondents to assess whether the adolescent in question has a problem, its chronicity, and whether this results in emotional distress, social impairment, or burden to the family. The psychometric properties of the SDQ tested in different cultural contexts and in clinical and community settings, are satisfactory (Goodman et al., 2000b; Goodman, 2001; Vostanis, 2006). Most SDQ studies report Cronbach's alpha coefficients above 0.70 for

the total Difficulties score for the three types of informants (van Widenfelt et al., 2003), and the lowest internal consistency for the Conduct (Koskelainen et al., 2001; Ronning et al., 2004) and Peer problems scales (Goodman et al., 2000a), which may indicate that these scales measure "more heterogeneous content than intended" (Palmieri and Smith, 2007, p. 190).

The SDQ has been used internationally and translated into more than 60 languages, which are readily available at http://www.sdqinfo.com/. The questionnaire took from 5 to 10 min to complete.

The Development and Well-Being Assessment inventory

The Development and Well-Being Assessment (DAWBA) was used as "gold standard" for testing the SDQ's concurrent and discriminant validity. It includes a package of questionnaires, interviews, and rating techniques used to generate ICD-10 and DSM-IV psychiatric diagnoses for children aged from 5 to 17. The specific disorders assessed were categorized into "internalizing," "externalizing," or "any mental disorder" (Goodman et al., 2000a).

Both mothers and adolescents responded to the DAWBA inventory, constructed for its administration in the community, which combines some of the best features of structured and semi-structured measures. When definite symptoms were identified by the structured questions, interviewers used open-ended questions, and supplementary prompts to get parents or the adolescent to describe the problem in their own words (Goodman et al., 2000a). On the basis of the comments of both mothers and adolescents recorded by the interviewers, a team of psychiatrists confirmed or rejected the preliminary computerized diagnoses and a final single diagnosis for each adolescent was thus obtained. The Hebrew translation was performed by the same procedure as described above for the SDQ.

SURVEY MODE

The survey used a face-to-face interview mode and was carried out at the respondents' homes between January 2004 and March 2005. Two trained interviewers interviewed the mother and adolescent simultaneously and independently. The mother was specifically targeted, as opposed to "any adult caretaker," because we assumed she is more frequently at home, more accessible and more aware of the health services used by the family members. As in other studies (Costello et al., 1996), we selected the mother as the adult respondent, unless she had not lived with the child for the 6 months preceding the interview.

A survey firm, Public Opinion and Marketing Research of Israel (PORI), employed interviewers and supervised the fieldwork, together with the staff of the Ministry of Health. The face-to-face mode was particularly important given the sensitive questions asked and the length of the interview. On average the mothers' interview took between 50 and 90 min and the adolescents' interview between 45 and 75 min, depending on the history of disorders of the adolescent.

Efforts were made to confirm the address and telephone number of families of all the adolescents included. An introductory letter explaining the objectives, randomization methods and confidentiality of the data and providing a phone number for possible queries was sent to each pre-designated target family. A week later, the interviewer arranged an interview date by phone and at the assigned date two interviewers made a home visit.

The mother was asked to sign a consent form for herself and her child and both were informed of their right to stop the interview at any time.

Mothers were interviewed in Hebrew ($N = 570$), Arabic ($N = 300$), or Russian ($N = 87$), according to the preference of respondents. Adolescents were interviewed in Hebrew ($N = 657$) and Arabic ($N = 300$), assuming that all adolescents who had immigrated from the Former Soviet Union were fluent in Hebrew by the time of the interview. The analyses presented here were made only on mothers and adolescents who responded in Hebrew. Among these, 553 mothers and 611 adolescents filled in the SDQ questionnaire.

TRAINING OF INTERVIEWERS

The 104 lay interviewers who participated in the data collection, mostly women, were college students and experienced survey interviewers. They were trained in small groups in an 8-h training session during which they went over the interview schedule, reviewed the questionnaires and were instructed regarding particular aspects of the different questions. After they completed the first three interviews, they went through the questionnaires with the fieldwork supervisor and reviewed any errors or omissions in each of the completed questionnaires.

CONFIDENTIALITY AND ETHICS

All interviewers signed a confidentiality form. All identifying information, except for the identity code, was stored separately from the questionnaire. Confidentiality could be infringed only for cases in which the adolescent reported sexual abuse or explicit suicidal intentions of which parents were unaware, cases that the law mandates should be conveyed to the appropriate authorities, though we found none. Parents signed consent forms for their own and their child's participation in the study, as approved by a Human Subjects Committee of the Schneider Medical Center for Children in Israel. Adolescents were explained the objectives and methods of the survey and could abstain from answering questions. Confidentiality was assured.

RESPONSE RATES

Table 1 shows response rates by gender and population groups. Overall response rate was 68.2%: 14.8% of the subjects were not located and 17.0% refused to participate in the study. Among the located subjects response rate approached 80%. Response rates varied among population groups with higher rates among boys than girls and higher among Druze, Muslim, and Christian respondents than among Jewish respondents. Also, there were higher response rates among adolescents living in mid-sized urban localities than among those living in larger cities.

Defining "respondents," "refusals," and "not contacted"

"Respondents." The subjects were classified as respondents, if the interviews of either mother or adolescent or both were performed. Twenty-two mothers refused to answer the questionnaire but signed and consented that their child participates, and 50 adolescent refused, although their mothers answered the questionnaire. In these cases, we based our diagnoses on one single source of information. Thus, we have 885 cases with two informants and 72 cases with only one informant (either adolescent or mother).

Table 1 | Fieldwork results by gender and population groups.

| | Total population | | Gender | | | | Type of locality | | | |
| | | | Male | | Female | | Jewish/mixed localities | | Arab localities | |
	N	%	N	%	N	%	N	%	N	%
Gross sample	1,670	100.0	854	100.0	816	100.0	1,370	100.0	300	100.0
Excluded:	268	16.0	139	16.3	130	15.9	267	19.5	1	0.3
(a) Ultra-orthodox Jews	245	14.7	132	15.5	114	14.0	245	17.9	0	0.0
(b) Abroad 1 year+	23	1.3	7	0.8	16	1.9	22	1.6	1	0.3
Net sample	1,402	100.0	715	100.0	686	100.0	1103	100.0	299	100.0
Not located	207	14.8	102	14.3	105	15.3	195	17.7	12	4.0
Refusals	238	17.0	116	16.2	121	17.6	219	19.9	19	6.4
Respondents	957	68.2	497	69.5	460	67.1	689	62.5	268	89.6
Of located: respondents	957	80.1	497	81.1	460	79.2	689	75.9	268	93.4

"Refusals." After sending an introductory letter, the family was contacted in order to convince both mother and adolescent to participate in the study. If the family refused, a second letter signed by the Mental Health Advisor to the Minister of Health (IL) was sent, appealing to them to contribute to the public good. If they still refused, the survey coordinator again tried to convince them to participate. Only after the third refusal, they were classified as "Refusals" ($N = 238$).

"Not located." If the introductory letter was returned by the post office, the interviewer visited the address to corroborate that the family had moved and to try to find out from neighbors the new address. If the interviewer could not obtain additional information, the school principal was approached and, with the help of an official letter from the Ministry of Education, was requested to help us find the adolescent's address and home telephone. If the child had left the school and the school authorities did not have any information, we approached the Ministry of Interior and tried to find out the new address of the family. If we failed, the case was considered "Not located" ($N = 207$).

INFLATION METHOD AND RESPONSE RATES

The sample was weighted back to the total population to compensate for unequal selection probabilities resulting from clustering effects and non-response. The weights were adjusted to make weighted sample totals conform to known population totals taken from reliable Central Bureau of Statistics (CBS) sources, after Jewish ultra-orthodox adolescents were removed. The weighted groups were chosen according to gender, age, and population groups. The categories were gender (male and female), age (14, 15, 16, 17 years), and population group (Jews and others born in Israel; Jews and others born abroad; Muslim and Christian Arabs; and Druze). The inflation method for each group of individuals was determined according to the known characteristics of the respondents and non-respondents in the given group. Response status was assigned according to the following categories: (a) respondent; (b) left the country; (c) refused to answer; (d) ultra-orthodox; (e) not located. Jewish ultra-orthodox youth were excluded from the poplation, as

explained above, as well as those who left the country. Respondents and refusals were included in the study population and those not contacted were proportionally divided into refusals and those who should not have been included in the study population from the start.

QUALITY CONTROL PROCEDURES

Data quality was controlled in a number of ways. Parents had to sign an informed consent form and, therefore, we could corroborate that the interview had indeed taken place. A few parents refused to sign the form out of fear that this would obligate them to something later on but agreed to carry out the interview. These parents were contacted again to make sure they had agreed to be interviewed. All questionnaires were reviewed upon reception by the field coordinator and missing data that could be provided by the interviewers were retrieved as soon as possible.

DATA ANALYSIS

Analyses on unweighted data were performed using the SPSS - 17 software package (SPSS Inc., Chicago, IL, USA). Internal consistency of the SDQ-H and distinct scales was assessed with Cronbach's alpha. An exploratory factor analysis (EFA) was performed to identify underlying factors and assess construct validity of the SDQ. This was followed by confirmatory factor analysis (CFA) in AMOS. Although construct validity is not synonymous with factor structure, we use here the term construct validity for consistency with Kline (1986) and Thompson (2004) and with the understanding that factor analysis is a commonly used way to assess the extent of construct validity.

Pearson product moment correlations between SDQ underlying scale constructs were calculated and *t*-tests were used comparing SDQ mean scores between DAWBA cases and non-cases (Farbstein et al., 2010) to test concurrent validity. Discriminant validity was assessed using receiver operating characteristic (ROC) curves, employing area under the curve (AUC) as an index of discriminant ability. Distributions of raw scores were used to determine the cut-off scores to identify normal, borderline, and abnormal bandings.

RESULTS OF THE FIELDWORK

Table 2 presents the sociodemographic characteristics of the survey participants who answered in Hebrew. The proportion of boys was slightly higher than that of girls and about 17% of adolescents lived with their divorced or single mother. The majority of respondents lived in families with up to three children. Almost half of the respondents' mothers had 13 years or more of education. Nearly 17% of the respondent' fathers did not work in paid employment; either due to a physical or a mental disability, because they were pensioners or still studying, or in prison or because they did not find work. Over 3/4 of the adolescents were born in Israel.

RESULTS

MEAN SCORES

Mean scores for the SDQ-H samples and comparisons by gender are presented in **Table 3**. The self-report was more gender-specific than the mothers' report. **Table 3** shows that mothers attributed higher mean scores to girls' regarding emotional symptoms and prosocial behaviors, while they attributed higher mean scores to boys regarding hyperactivity. Among adolescents, the findings show that girls attributed themselves higher scores in Total Difficulties, Conduct problems, Hyperactivity, Prosocial behaviors, and Impact, while boys attributed themselves higher scores in the Emotional symptoms scale.

We compared mean scores of mothers' ratings with self ratings and found that they were significantly different for all scales (data not in table). Mothers attributed higher Prosocial scores to their child than did the adolescents themselves ($F = 5.114$; $p < 0.001$), while adolescents rated themselves with significantly higher scores

than their mothers regarding Total Difficulties ($F = 5.119$; $p < 0.001$), Emotional Symptoms ($F = 6.731$; $p < 0.001$), Conduct problems ($F = 13.534$; $p < 0.001$), Hyperactivity ($F = 9.935$; $p < 0.001$), and Peer problems ($F = 7.257$; $p < 0.001$).

INTERNAL CONSISTENCY

The Total Difficulties scale had acceptable internal consistency for the parent- ($\alpha = 0.76$) and self-versions ($\alpha = 0.72$). For both versions, the Hyperactivity scale was the most reliable ($\alpha = 0.73$ and 0.71, respectively), and the Peer problem scale the least ($\alpha = 0.35$ and 0.22, respectively) According to the parent and the self-versions, the α coefficients for the Emotional Symptoms scale were 0.56 and 0.63, for the Conduct Problems 0.48 and 0.51, and for the Prosocial Behavior scale 0.53 and 0.56, respectively.

CONSTRUCT VALIDITY

Exploratory factor analysis was performed on the SDQ-H mother-version (**Figure 1**) and on the self-version (**Figure 2**). In the current EFA analysis principal components analysis (PCA) was used as the algorithm to maximize the variance explained (Gorsuch, 1993). Varimax rotation was applied as it maximizes the squared loadings of a factor, gives similar results to many oblique factorial solutions and produces an orthogonal solution (see Gorsuch, 1993). To identify the number of factors a scree plot was generated of PCA of the actual SDQ data superimposed on data values from 20 (based on previous recommendations) simulated datasets of

Table 2 | Socio-demographic characteristics of adolescents responding in Hebrew (raw numbers and weighted proportions).

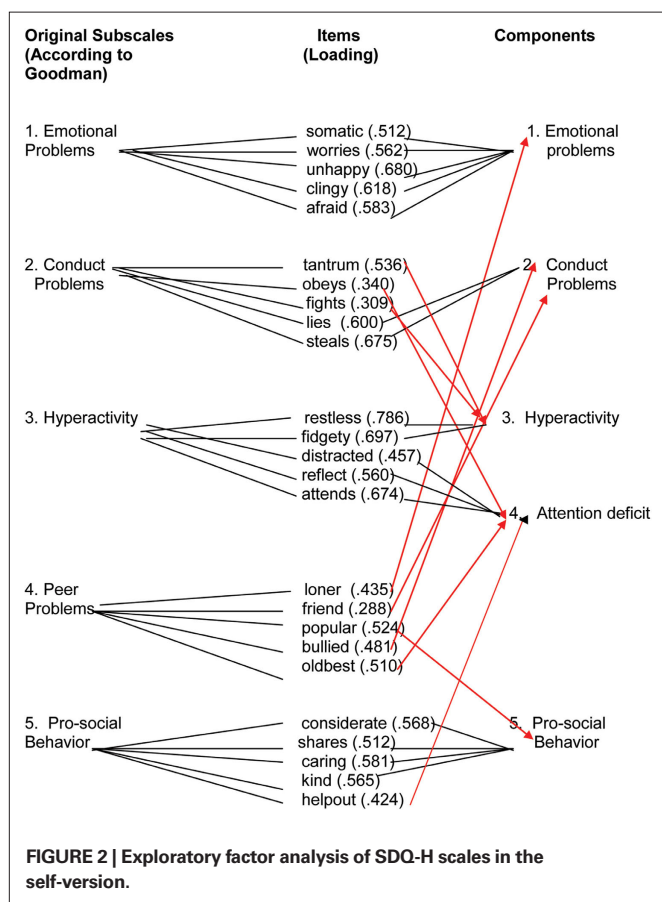| Characteristic	N	%
GENDER		
Boys	348	51.2
Girls	309	48.8
MARITAL STATUS OF PARENTS		
Married	538	83.5
Single/divorced/widowed	103	16.5
NUMBER OF CHILDREN IN FAMILY		
1	23	3.9
2	143	24.6
3	209	31.6
4–6	217	32.9
7–16	49	7.1
MATERNAL YEARS OF SCHOOLING		
0–11	93	14.2
12	231	36.9
13 and over	297	48.8
EMPLOYMENT STATUS OF FATHER		
Employed	497	83.1
Unemployed	101	16.9
COUNTRY OF ORIGIN		
Israel	528	76.3
Other	129	23.7
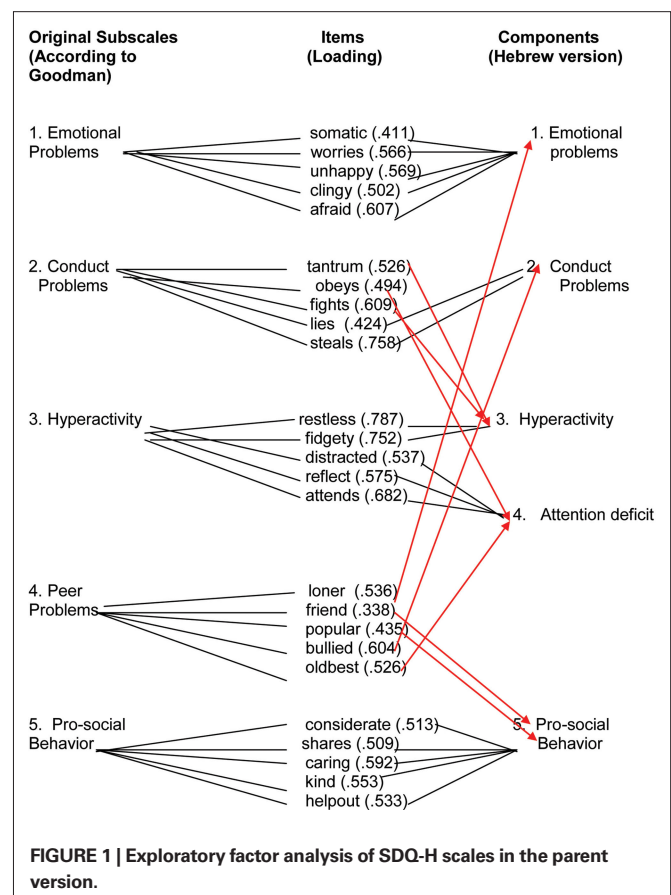
FIGURE 1 | Exploratory factor analysis of SDQ-H scales in the parent version.

Table 3 | Mean (SD) scores for SDQ-Hebrew for the parent and self reported versions by gender.

SDQ scales	Parent-report					Self-report				
	Male (N = 298)	Female (N = 294)	Total sample (N = 533)	F	p	Male (N = 317)	Female (N = 294)	Total sample (N = 611)	F	p
Total difficulties	6.72 (4.8)	6.39 (4.7)	6.57 (4.7)	0.700	ns	8.23 (4.5)	9.29 (4.6)	8.84 (4.6)	9.408	<0.01
Emotional symptoms	1.56 (1.6)	1.95 (1.8)	1.74 (1.7)	7.222	<0.01	2.09 (1.8)	1.92 (1.8)	2.65 (2.0)	42.158	<0.001
Conduct problems	1.48 (1.5)	1.38 (1.4)	1.44 (1.4)	0.716	ns	1.75 (1.6)	2.04 (1.5)	1.75 (1.5)	0.037	ns
Hyperactivity	2.19 (2.1)	1.75 (2.0)	1.99 (2.1)	6.135	<0.05	2.91 (2.2)	2.99 (1.9)	2.94 (2.2)	0.000	ns
Peer problems	1.49 (1.5)	1.31 (1.4)	1.41 (1.4)	2.222	ns	1.48 (1.3)	2.34 (1.6)	1.51 (1.3)	0.796	ns
Prosocial behavior	8.47 (1.6)	8.75 (1.5)	8.60 (1.5)	4.61	<0.05	8.00 (1.6)	8.28 (1.6)	8.24 (1.6)	14.315	<0.001
Impact	0.35 (1.2)	0.33 (1.2)	0.34 (1.2)	0.039	ns	0.18 (0.7)	0.25 (1.0)	0.20 (0.8)	6.275	<0.05

FIGURE 2 | Exploratory factor analysis of SDQ-H scales in the self-version.

equal row and column size to the actual dataset (Gorsuch, 1993). **Figure 3** depicts the eigenvalues of the SDQ data plotted against the simulated eigenvalues. This indicates that the eigenvalues for the first five components in the data exceed the simulated eigenvalues, thereby providing further support for the five component solution. To guide interpretation of the results of EFA, values in excess of 0.3 were used as is common for self-reported data, and for CFA statistical significance model fit and significant loadings were used.

For the mother SDQ-H version (**Figure 1**) all items in the Emotional Symptoms and Prosocial Behavior scales loaded on the corresponding factors. The Hyperactivity-inattention scale was split into two distinct factors: "restless" and "fidgeting" loaded on the hyperactivity factor, while "distracted," "thinks things out," and "attention span" loaded on the attention deficit factor. Additional items, originally belonging to other scales, also loaded with the hyperactivity (i.e., "tantrums" and "fights") or the attention deficit (i.e., "obedient") factors. The Peer problems scale did not retain its original structure, as its items loaded on other factors.

Also for the self-version the Emotional problems factor retained its original structure (**Figure 2**). The Hyperactivity scale was also here divided into two factors, one representing the more external behaviors and the other the attention elements. The Peer problems scale was cross-loaded to all the other factors and did not retain its original structure. Three of the items in the conduct scale were cross-loaded to the Hyperactivity scale.

Confirmatory factor analysis was performed to estimate the goodness of fit between the SDQ structure and the sample data. Five factors were specified. Loadings on all items were significant and no errors were set to correlate and the *post hoc* improvement indices were not used. Goodness of fit was examined using the χ^2 and RMSEA. Chi squares were examined where significant values mean that the model specified significantly deviates from the data (i.e., misfit). RMSEA was chosen since it rewards model parsimony and is generally insensitive to sample size and, unlike most fit indices, confidence intervals around the point estimate are available. Good model fit is indicated by an RMSEA value under 0.06 and values below 0.08 indicate moderate fit. Lower values reflect a better fit to the model. Results show that a five-factor model was a reasonable fit to the data for the parent version ($\chi^2 = 1096.71$, df = 289, $p < 0.01$; RMSEA = 0.07, 90% CI 0.066–0.075), and a good fit for the self-version ($\chi^2 = 612.413$, df = 269, $p < 0.01$; RMSEA = 0.047, 90% CI 0.042–0.052). The final model consisted of four moderately to highly correlated factors: emotional, conduct, hyperactivity, and peer related problems. A fifth and distinct prosocial factor was negatively related to the four problem factors.

CONSTRUCT DISTINCTNESS

Pearson correlations between the distinct SDQ-H dimension scores showed, for the parent version, a positive association of conduct problems with hyperactivity ($r = 0.54$, $p < 0.001$) and emotional symptoms ($r = 0.40$, $p < 0.01$) scores. Although, emotional

symptoms scores were associated significantly and positively with peer problems and hyperactivity scores ($r = 0.35$ and 0.27, $p < 0.01$; respectively), and between peer- and conduct problems scores ($r = 0.23$, all $p < 0.05$), the associations were small in magnitude (see **Table 4**). For the self-version (in the parentheses), Pearson correlations show very similar results in magnitude, direction and significance.

CONCURRENT VALIDITY

Concurrent validity was assessed against independent psychiatrists' diagnoses according to DAWBA (Farbstein et al., 2010). **Table 5** shows that mean scores for the mother- and self-versions were significantly higher for adolescents with an internalizing disorder than for those without, on all scales ($p < 0.001$), except for the Prosocial scale. Mean scores for the mother's and self-reports on the Conduct problems and Hyperactivity scales were higher for those with an externalizing disorder. The Total Difficulties score and all the SDQ scales showed significantly higher scores when any mental disorder was present, except for the Prosocial self-report.

FIGURE 3 | Twenty simulated random eigenvalues superimposed on those in the SDQ data.
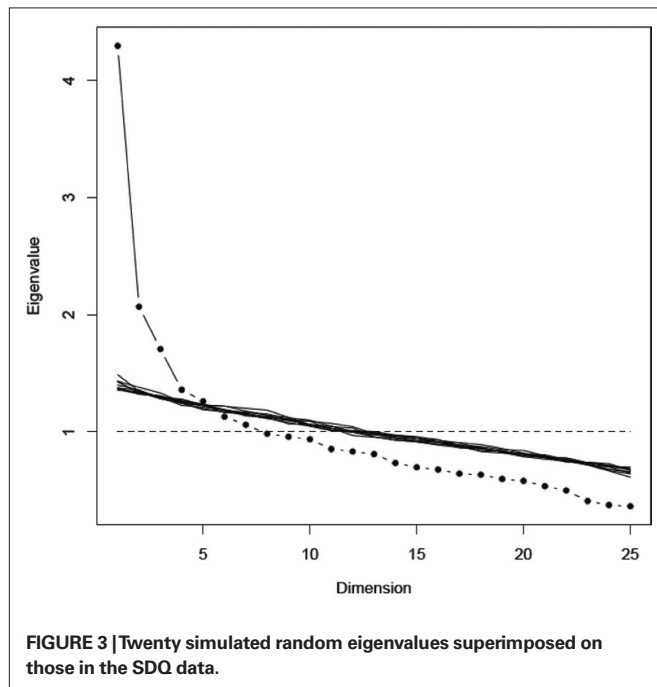
DISCRIMINANT VALIDITY

We compared adolescents with and without a mental disorder according to DAWBA, using ROC analysis, where sensitivity and specificity are combined and provide an AUC, with values varying between 0 and 1 (**Table 6**). Yasong et al. (2008), who made in China a study similar to ours, claim that a score of <0.6 suggests that discrimination is no better than chance; between 0.6 and 0.75 suggests it is fair; between 0.75 and 0.90, that the discrimination is good; and above 0.90, that it is very good. We found that all the scales, except the Prosocial Behavior scale, discriminated fairly between those with and without any mental disorder. The discriminant power of the Emotional Symptoms and Peer problems scales in the self-reported version vis-à-vis internalizing disorder were good and fair (0.84 and 0.65, respectively), but they had only fair discriminant power in the parent version. The discriminant power of the Conduct problems and the Hyperactivity scales in the parent version vis-à-vis externalizing disorders were good and very good (0.77 and 0.91, respectively), whereas in the self-reported version they were fair (0.73 and 0.75, respectively).

NORMATIVE SCORES

We used Goodman's trichotomy for the SDQ screening (Goodman et al., 2004), according to which about 10% of the population should be classified as "probable cases," another 10% as "possible cases," and the remaining 80% as "unlikely cases." The cut-off points were contextually based, according to the prevalence of any mental disorder in this age group in the country of concern. In Israel the prevalence of any mental disorder among 14- to 17-year-olds is 12% (Farbstein et al., 2010), therefore the top band was defined as the top 10–12%. The next band down was chosen to be the same size, and the rest was considered to be in the "unlikely" category. Given that we had a limited sample size, the observed percentage of adolescent within each range does not conform exactly to the expected percentage. It is to be expected that with a large sample size, the cutting points could be closer to the 10%- 10%- 80% proportions. **Table 7** shows the normative scores for detecting "caseness" for both informant versions of the SDQ and the percentage of adolescents falling within each category. No significant differences in the percentage of probable diagnosis by gender were found neither for the parent or the self-versions. These norms were not different for the Israeli population answering in Hebrew than for the total Israeli sample.

Table 4 | Correlation matrix for the SDQ-H scales for parent and (self) reports.

Scales	Prosocial	Peer problems	Hyperactivity	Emotional problems	Conduct problems	Total difficulties
Prosocial	1					
Peer problems	−0.228** (−0.135**)	1				
Hyperactivity	−0.314** (−0.201**)	0.183** (0.129**)	1			
Emotional problems	−0.067 (0.037)	0.350** (0.269**)	0.271** (0.167**)	1		
Conduct problems	−0.358** (−0.276**)	0.230** (0.108*)	0.539** (0.432**)	0.395** (0.321**)	1	
Total Difficulties	−0.342** (−0.206**)	0.582** (0.491**)	0.760** (0.716**)	0.708** (0.694**)	0.756** (0.695**)	1

*$p < 0.05$, **$p < 0.01$.

Table 5 | Concurrent validity of the SDQ-H scores according to DAWBA diagnosis of internalizing, externalizing and any disorders.

Score	Internalizing			Externalizing			Any disorder****		
	Present	Absent	t-value	Present	Absent	t-value	Present	Absent	t-value
PARENT-REPORT									
Total Difficulties	10.2 ± 5.8	6.5 ± 4.1	6.53***	12.4 ± 4.8	6.6 ± 4.6	6.66***	11.3 ± 5.5	5.9 ± 4.1	10.03***
Emotional symptoms	3.5 ± 2.5	1.6 ± 1.5	9.54***	1.8 ± 1.3	1.7 ± 1.8	0.32	3.0 ± 2.1	1.6 ± 1.5	6.86***
Conduct problems	1.9 ± 1.4	1.4 ± 1.4	3.24**	3.0 ± 1.8	1.4 ± 1.4	6.20***	2.6 ± 1.7	1.3 ± 1.3	7.52***
Hyperactivity	2.5 ± 2.3	2.0 ± 2.1	2.13*	6.1 ± 2.4	1.9 ± 1.9	11.40***	3.8 ± 2.7	1.7 ± 1.8	8.59***
Peer problems	2.2 ± 1.6	1.6 ± 1.4	3.52***	1.5 ± 1.4	1.6 ± 1.5	0.69	2.0 ± 1.7	1.3 ± 1.4	3.62***
Prosocial behavior	8.4 ± 2.1	8.7 ± 1.6	1.45	7.6 ± 2.1	8.7 ± 1.6	3.85***	8.2 ± 1.9	8.7 ± 1.5	2.72**
Impact	1.5 ± 2.6	0.2 ± 0.9	8.69***	1.2 ± 1.4	0.3 ± 1.2	4.02***	1.6 ± 2.4	0.1 ± 0.7	11.26***
SELF-REPORT									
Total Difficulties	14.4 ± 4.4	8.6 ± 4.4	10.36***	11.9 ± 4.7	9.0 ± 4.6	3.07**	13.6 ± 4.9	8.2 ± 4.2	10.88***
Emotional symptoms	5.2 ± 2.1	2.4 ± 1.9	11.82***	2.1 ± 1.8	2.6 ± 2.1	1.21	4.1 ± 2.6	2.4 ± 1.8	6.94***
Conduct problems	2.6 ± 1.5	1.8 ± 1.5	4.76***	3.3 ± 2.1	1.8 ± 1.5	5.13***	2.9 ± 1.8	1.6 ± 1.4	8.11***
Hyperactivity	4.0 ± 1.9	2.8 ± 2.1	4.49***	4.9 ± 2.4	2.8 ± 2.0	5.03***	4.5 ± 2.2	2.7 ± 2.1	7.04***
Peer problems	2.6 ± 1.6	1.7 ± 1.4	4.57***	1.6 ± 1.2	1.8 ± 1.5	0.68	2.1 ± 1.6	1.4 ± 1.3	4.76***
Prosocial behavior	8.3 ± 1.5	8.3 ± 1.5	0.12	7.7 ± 1.5	8.3 ± 1.5	1.95	8.0 ± 1.5	8.3 ± 1.6	1.62
Impact	1.5 ± 2.1	0.2 ± 0.7	11.82***	0.4 ± 1.0	0.3 ± 0.9	0.75	1.2 ± 1.9	0.2 ± 0.6	9.48***

Mean scores ± SD are shown.
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.
****Internalizing or externalizing disorders or both.

DISCUSSION

This study shows that, in general, SDQ-H has acceptable to good internal consistency, and construct, concurrent, and discriminant validity. The Hyperactivity and Emotional symptoms scales were the most robust, while the Peer problems scale was the weakest. The Total Difficulties score for both parent and self-versions had good reliability and discriminant validity. These results are important to determine the viability of using SDQ-H as a screening measure in community surveys and as an evaluation measure for clinical practice.

COMPARABILITY OF MEAN SDQ-H SCORES

Similarly to the Dutch (van Widenfelt et al., 2003), our mean scores were lower than the British on the Emotional Symptoms and Hyperactivity scales and Total Difficulties scores. This variation could reflect age differences, lower expectations and higher tolerance for misconduct, or more lenient judgment of problematic behaviors and traits by parents in the different populations (Marzocchi et al., 2004).

Unlike the French (Shojaei et al., 2009), but consistent with the Dutch (Muris et al., 2003) and Finnish (Koskelainen et al., 2001) studies, girls scored higher on both informant versions of the Total Difficulties scale. The discrepancy between the studies may be attributed to age grouping, as the French study included younger children, while our sample included senior adolescents, among whom the likelihood of internalizing disorders among girls is higher (Farbstein et al., 2010). Consistently with most studies, mothers scored their boys higher on the Hyperactivity scale and their girls higher on Prosocial Behavior. The self-version revealed no gender differences.

Other studies (Koskelainen et al., 2001) have also found a tendency for youth to rate more symptoms than parents, noticing that, compared to their parents' or teachers', children and adolescents in non-clinical settings describe themselves as having more problems, while those in a clinical sample are prone to under-report their mental problems (Becker et al., 2004).

INTERNAL CONSISTENCY

In both informant versions, the Total Difficulties and Hyperactivity scale showed acceptable internal consistency, while the Emotional symptoms, Conduct problems and Prosocial behavior scales had a somewhat lower internal consistency. The Peer problems scale had an unacceptable reliability. The Peer problems' known low reliability (Goodman, 2001; Shojaei et al., 2009), has been attributed to the possibility that it measures a heterogeneous construct (Palmieri and Smith, 2007). The disappearance of the Peer problems scale as a factor in our EFA, added to its low internal consistency, supports this explanation.

CONSTRUCT VALIDITY

Our findings suggest that, except for the Peer problems which cross-loaded with all other factors, and to a lesser extent also the Conduct problems of which three items cross-loaded with the hyperactivity factor, the factors underlying SDQ-H scales are distinct constructs. In accord with studies conducted in different settings and languages (Hawes and Dadds, 2004; Woerner et al., 2004b), we found that the final model consisted of four moderately to highly correlated factors comprising the Total Difficulties score, and a fifth and distinct prosocial factor inversely related to the four problem factors. This five-factor model was a reasonable fit to the data for both SDQ-H versions.

The CFA shows that the emotional and prosocial factors retained their original structure whereas the hyperactivity-inattention scale split into two distinct factors. Cross-loading of the "obedience" item

Table 6 | Ability of SDQ scores to distinguish between adolescents with and without internalizing, externalizing or any disorder, according to DAWBA clinical diagnosis.

Type of disorder	Scale	Area under the curve	Asymptotic 95% confidence interval	
			Lower bound	Upper bound
ANY DISORDER				
Parent	Total Difficulties score	0.75	0.70	0.81
	Emotional symptoms	0.71	0.65	0.76
	Conduct problems	0.69	0.64	0.75
	Hyperactivity	0.68	0.62	0.74
	Peer problems	0.58	0.52	0.64
	Prosocial behavior	0.43	0.37	0.50
Self	Total Difficulties score	0.82	0.77	0.86
	Emotional symptoms	0.75	0.69	0.80
	Conduct problems	0.73	0.67	0.78
	Hyperactivity	0.71	0.66	0.76
	Peer problems	0.62	0.56	0.68
	Prosocial behavior	0.46	0.40	0.51
INTERNALIZING DISORDERS				
Parent	Total Difficulties score	0.69	0.62	0.75
	Emotional symptoms	0.74	0.68	0.80
	Conduct problems	0.62	0.55	0.69
	Hyperactivity	0.57	0.50	0.64
	Peer problems	0.62	0.54	0.69
	Prosocial behavior	0.49	0.41	0.57
Self	Total Difficulties score	0.82	0.78	0.87
	Emotional symptoms	0.84	0.79	0.89
	Conduct problems	0.68	0.62	0.74
	Hyperactivity	0.66	0.60	0.72
	Peer problems	0.65	0.58	0.71
	Prosocial behavior	0.50	0.43	0.57
EXTERNALIZING DISORDERS				
Parent	Total Difficulties score	0.84	0.79	0.90
	Emotional symptoms	0.58	0.48	0.67
	Conduct problems	0.77	0.69	0.86
	Hyperactivity	0.91	0.87	0.96
	Peer problems	0.45	0.34	0.57
	Prosocial behavior	0.35	0.24	0.47
Self	Total Difficulties score	0.68	0.58	0.79
	Emotional symptoms	0.44	0.33	0.55
	Conduct problems	0.73	0.62	0.84
	Hyperactivity	0.75	0.65	0.85
	Peer problems	0.48	0.37	0.58
	Prosocial behavior	0.38	0.28	0.48

to the inattention factor may mean that, in this population, disobedience is not attributed to the adolescent's will to break a norm but rather to his/her inability to pay attention and understand what is being demanded. It has been argued that this item could have a different function or meaning among Middle Eastern than among Western children (Thabet et al., 2000), though studies carried out in

Western countries that consistently support the original five-factor structure (Goodman, 2001; van Widenfelt et al., 2003), have also found limited associations between the "obedient" item and the conduct problems factor, with which it was originally associated.

The low internal consistency of the Peer problem scale may also affect attempts to replicate previous factor analyses, since "unreliable items … can … lead to different factor structures" (Ozer et al., 2009, p. 918). A study conducted in China (Yasong et al., 2008), found low validity for the Conduct and Peer problems scales and attributed this to greater cross-cultural acceptance and consistency of how prosocial behavior, emotional disorders, and hyperactivity should be expressed than about the types of behaviors indicative of positive peer relationships or of oppositional and conduct problems.

CONSTRUCT DISTINCTNESS
Correlations between SDQ scales scores were small to moderate in magnitude, which indicates their distinctness. The Total Difficulties score was highly correlated with the Hyperactivity, Emotional problems, and Conduct problems scales and this indicates a large contribution of these scales to the summary scale.

CONCURRENT VALIDITY
The SDQ Hyperactivity and Conduct problems scales were significantly associated with externalizing disorders according to DAWBA. Regarding internalizing disorders, all four problem scales differentiated well, even when excluding adolescents with co-morbidity. In sum, the correlations between SDQ scales' predictions of psychopathology and the diagnoses made with DAWBA showed sound external validity.

DISCRIMINANT VALIDITY
As expected,(Goodman, 2001; Goodman et al., 2004), ROC analyses showed that parents discriminated better than their children regarding externalizing disorders, while adolescents discriminated better regarding their internalizing disorders (Van Roy et al., 2008). Mothers had good discrimination for Conduct problems scale and very good discrimination for Hyperactivity scale vis-à-vis externalizing disorders. Adolescents showed good discrimination for the Emotional symptoms scale and fair discrimination for the Peer problems scale vis-à-vis internalizing disorders.

Although better predictions are obtained when the multi-informant algorithm, based on both symptoms and impact scores is used since the total error rate is reduced (Goodman et al., 2004), our results, based on single sources, support the SDQ-H's power to discriminate between adolescents with and without a mental disorder.

GENERAL COMMENTS
Comparing self and mothers' reports
Given the SDQ's widespread use in the community and in clinical settings, an important question is whether information gathered only from adolescents is as reliable and valid as data received from their parents. We found that the adolescents' version had as good internal consistency as the mothers' and even better construct validity. Total Difficulties showed better discriminant validity for the

Table 7 | Normative scores for detecting "caseness" level for the SDQ parent and self-report versions and percentage of respondents.

SDQ	Normal range		Borderline range		Abnormal range	
	Raw score	%	Raw score	%	Raw score	%
PARENT-REPORT						
Total Difficulties	0–9	75.9	10–12	11.7	13–40	12.4
Emotional symptoms	0–2	74.7	3	10.2	4–10	15.1
Conduct problems	0–2	81.0	3	11.2	4–10	7.8
Hyperactivity	0–3	78.2	5	9.4	5–10	12.4
Peer problems	0–2	75.7	3	12.3	4–10	12.0
Prosocial behavior	8–10	80.2	7	9.7	0–6	10.0
SELF-REPORT						
Total Difficulties	0–12	78.2	13–14	9.0	15–40	12.8
Emotional symptoms	0–4	81.5	5	8.0	6–10	10.5
Conduct problems	0–2	71.8	3	14.4	4–10	13.8
Hyperactivity	0–4	78.2	5	11.0	6–10	10.8
Peer problems	0–2	73.2	3	14.3	4–10	12.5
Prosocial behavior	8–10	72.9	7	14.3	0–6	13.4

adolescents' than the mothers' version for internalizing disorders but the opposite for externalizing disorders. Satisfactory reliability and good construct and concurrent validity were found for both versions. CFA showed better fit for the adolescents' than for the parent-report. Concurrent validity equally discriminated between disorders in the SDQ-H parent and self-reports.

The significant contribution of the self-version to diagnostic status, particularly if there is no other source of information, has been reported (Becker et al., 2004). Although ADHD should be diagnosed when symptoms are present in two or more settings and thus the information provided by mothers and teachers is necessary (Goodman et al., 2004), we found good discriminant scores in the Hyperactivity scale for the self reports. We conclude that adolescents' self-ratings contribute to the diagnosis and the use of the self-report version when no other possibility exists is valid.

Cultural, semantic, and language differences
Questions beyond sample size and age examined arise in the case of SDQ translations carried out in cultures markedly different to that where the SDQ originated. Three possible explanations for the incomplete agreement between our EFA and the original SDQ structure are: (a) there are inherent problems in the scale, which measures more heterogeneous content than intended (Palmieri and Smith, 2007); (b) different cultural norms, social desirability and parental expectations produce different results (Marzocchi et al., 2004); (c) the translation suffers from lack of semantic equivalence (Flaherty et al., 1988). A French study claims that the item "bullied" was rated "partly true" more frequently by French than UK parents and that the "difference could be explained by the translation of the corresponding item rather than the transcultural differences, given the French translation of that item has a milder meaning than the corresponding English item" (Shojaei et al., 2009, p. 745).

Although a comparative study found that a predictive algorithm developed for use in England worked well in Bangladesh despite the difference in language, culture and socio-economic circumstances (Goodman et al., 2000a), the question remains whether the algorithm, which can predict diagnoses in clinic samples with a good degree of accuracy, is equally useful when applied to community samples.

LIMITATIONS
The sample size might be relatively small for certain analyses. EFA requires a large sample before it settles into a reliable pattern and the normative data for the different subscales and versions require large sample sizes in order to present very distinct cut-off points. Another potential limitation is that analyses were performed independently on mothers and adolescents and therefore we could not test in this population whether SDQ predictions work best when they are multi-informant (Goodman et al., 2004).

Regarding concurrent validity, which used the DAWBA as the gold standard to assess the SDQ-H, there is a "small potential for circularity" (Goodman and Goodman, 2010, p. 7) between the SDQ and the DAWBA because high SDQ scores may lead to administering in full some DAWBA sections which do not screen positive on the DAWBA's own screening questions. "Collecting this additional DAWBA information is occasionally the basis for assigning diagnoses which would otherwise have been missed. This cannot explain the results observed, however, as a strong association with prevalence remained after excluding the mean scores of children with a disorder."(Goodman and Goodman, 2010, p. 7)

CONCLUSION
The psychometric validation of the SDQ in a non-Western cultural context and language contributes to the general efforts to better understand the value of standardized evaluation measures in clinical practice and community settings.

Both parent- and self-reports seem to discriminate fairly well between cases and non-cases, as shown by the fact that mean SDQ-H scores were significantly higher for adolescents with an internalizing disorder than for those without, on all scales, and mean scores on the Conduct problems and Hyperactivity scales

were higher for those with an externalizing disorder. The Total Difficulties score and all the SDQ scales showed significantly higher scores when any mental disorder was present.

Regarding reliability, although scores vary for the different SDQ-H scales, the Total Difficulties and the Hyperactivity scale for both mothers and adolescents had adequate alpha scores, above the 0.7 threshold which indicates acceptable internal consistency. Cronbach's alphas of 0.6 imply 60% true score and 40% error. Accordingly, 0.6 is reasonable for a screening measurement at best. Therefore, we attribute acceptable reliability to the Emotional symptoms scale for the self-version. The Peer problems and the Conduct problems scales, as in other studies, showed the lowest reliability (Goodman et al., 2000a; Koskelainen et al., 2001; Ronning et al., 2004) and the factor analyses also showed they some of the items belonging to these scales cross-loaded with other factors.

The current findings point to reasonable SDQ-H psychometric properties and we believe they are sufficiently promising to warrant further evaluation of SDQ-H as a screening tool. Given the fact that

parents are only partially aware of their children's internalizing disorders and the adolescents are only partially aware of their own externalizing disorders, as shown with the ROC analyses, we highly recommend a multiple-informant approach that also includes teachers' ratings.

REFERENCES
Achenbach, T. M., and Howell, C. T. (1993). Are American children's problems getting worse? A 13-year comparison. *J. Am. Acad. Child Adolesc. Psychiatry* 32, 1145–1154.

Alyahri, A., and Goodman, R. (2006). Validation of the Arabic Strengths and Difficulties Questionnaire and the Development and Well-Being Assessment. *La Revue Sante Mediterr. Orientale* 12, S138–S146.

Becker, A., Hagenberg, N., Roessner, V., Woerner, W., and Rothenberger, A. (2004). Evaluation of the self-reported SDQ in a clinical setting: do self-reports tell us more than ratings by adult informants? *Eur. Child. Adolesc. Psychiatry* 13, II17–II24.

Canino, G., Shrout, P. E., Rubio-Stipec, M., Bird, H. R., Bravo, M., Ramirez, R., Chavez, L., Alegría, M., Bauermeister, J. J., Hohmann, A., Ribera, J., García, P., and Martínez-Taboas, A. (2004). The DSM-IV rates of child and adolescent disorders in Puerto Rico: prevalence, correlates, service use, and the effects of impairment. *Arch. Gen. Psychiatry* 61, 85–93.

Costello, E. J., Angold, A., Burns, B. J., Stangl, D. K., Twee, D. L., Erkanli, A., and Worthman, C. M. (1996). The Great Smoky Mountains Study of Youth: goals, designs, methods, and the prevalence of DSM-III-R disorders. *Arch. Gen. Psychiatry* 53, 1129–1136.

Costello, E. J., Farmer, E. M., Angold, A., Burns, B. J., and Erkanli, A. (1997). Psychiatric disorders among American Indian and white youth in Appalachia: the Great Smoky Mountains Study. *Am. J. Public Health* 87, 827–832.

Farbstein, I., Mansbach-Kleinfeld, I., Levinson, D., Goodman, R., Levav, I., Vograft, I., Kanaaneh, R., Ponizovsky, A. M., Brent, D. A., and Apter, A. (2010). Prevalence and correlates of mental disorders in Israeli adolescents: results from a national mental health survey. *J. Child Psychol. Psychiatry* 51, 630–639.

Flaherty, J. A., Gaviria, F. M., Pathak, D., Mitchell, T., Wintrob, R., Richman, J. A., and Birz, S. (1988). Developing instruments for cross-cultural psychiatric research. *J. Nerv. Ment. Dis.* 176, 257–263.

Goodman, A., and Goodman, R. (2009). Strengths and Difficulties Questionnaire as a dimensional measure of child mental health. *J. Am. Acad. Child Adolesc. Psychiatry* 48, 400–403.

Goodman, A., and Goodman, R. (2010). Population mean scores predict child mental disorder rates: validating SDQ prevalence estimators in Britain. *J. Child Psychol. Psychiatry.* doi: 10.1111/j.1469-7610.2010.02278.x.

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.

Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *J. Am. Acad. Child Adolesc. Psychiatry* 40, 1337–1345.

Goodman, R., Ford, T., Corbin, T., and Meltzer, H. (2004). Using the Strengths and Difficulties Questionnaire (SDQ) multi-informant algorithm to screen looked-after children for psychiatric disorders. *Eur. Child. Adolesc. Psychiatry* 13, II25–II31.

Goodman, R., Ford, T., Richards, H., Gatward, R., and Meltzer, H. (2000a). The Development and Well-Being Assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *J. Child Psychol. Psychiatry* 41, 645–655.

Goodman, R., Renfrew, D., and Mullick, M. (2000b). Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *Eur. Child. Adolesc. Psychiatry* 9, 129–134.

Goodman, R., and Scott, S. (1999). Comparing the strengths and difficulties questionnaire and the child behavior checklist: is small beautiful? *J. Abnorm. Child Psychol.* 27, 17–24.

Gorsuch, R. L. (1993). *Factor Analysis*, 2nd Edn. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hawes, D. J., and Dadds, M. R. (2004). Australian data and psychometric properties of the Strengths and Difficulties Questionnaire. *Aust. N. Z. J. Psychiatry* 38, 644–651.

Heiervang, E., Goodman, A., and Goodman, R. (2008). The Nordic advantage in child mental health: separating health differences from reporting style in a cross-cultural comparison of Psychopathology. *J. Child Psychol. Psychiatry* 49, 678–685.

Heiervang, E., Stormark, K. M., Lundervold, A. J., Heimann, M., Goodman, R., Posserud, M. B., Ullebo, A. K., Plessen, K. J., Bjelland, I., Lie, S. A., and Gillberg, C. (2007). Psychiatric disorders in Norwegian 8- to 10-year-olds: an epidemiological survey of prevalence, risk factors, and service use. *J. Am. Acad. Child Adolesc. Psychiatry* 46, 438–447.

Kline, P. (1986). *A Handbook of Test Construction: Introduction to Psychometric Design.* New York: Routledge.

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

Mansbach-Kleinfeld, I., Levinson, D., Farbstein, I., Apter, A., Levav, I., Kanaaneh, R., Stein, N., Erhard, R., Palti, H., Khwaled, R., and Ponizovsky, A. M. (2010). The Israel Survey of Mental Health among Adolescents: aims, methods, strengths and limitations. *Isr. J. Psychiatry Relat. Sci.* 2, 8–17.

Marzocchi, G., Capron, C., Di Pietro, M., Tauleria, E. D., Duyme, M., Frigerio, A., Gaspar, M. F., Hamilton, H., Pithon, G., Simoes, A., and Theroud, C. (2004). The use of the Strengths and Difficulties Questionnaire (SDQ) in Southern European countries. *Eur. Child. Adolesc. Psychiatry* 13(Suppl. 2), 40–46.

Meltzer, H., Gatward, R., Goodman, R., and Ford, T. (2003). Mental health of children and adolescents in Great Britain. *Int. Rev. Psychiatry* 15, 185–187.

Muris, P., Meesters, C., and Van Den Berg, F. (2003). The Strengths and Difficulties Questionnaire (SDQ): further evidence for its reliability and validity in a community sample of Dutch children and adolescents. *Eur. Child. Adolesc. Psychiatry* 12, 1–8.

Ozer, Z. C., Fiorat, M. Z., and Bektas, H. A. (2009). Confirmatory and exploratory factor analysis of the caregiver quality of life index-cancer with Turkish samples. *Qual. Life Res.* 18, 913–921.

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

Ronning, J. A., Handegaard, B. H, Sourander, A., and Morch, W. T.

(2004). The Strengths and Difficulties Self-Report Questionnaire as a screening instrument in Norwegian community samples. *Eur. Child Adolesc. Psychiatry* 13, 73–82.

Shojaei, T., Wazana, A., Pitrou, I., and Kovess V. (2009). The strengths and difficulties questionnaire: validation study in French school-aged children and cross-cultural comparisons. *Soc. Psychiatry Psychiatr. Epidemiol.* 44, 740–747.

Surgeon General. (1999). *Mental Health: A Report of the Surgeon General.* Available at: http://mentalhealth.samhsa.gov/cmhs/surgeongeneral/surgeongeneralrpt.asp

Thabet, A. A., Stretch, D., and Vostanis, P. (2000). Child mental health problems in Arab children: application of the Strengths and Difficulties Questionnaire. *Int. J. Soc. Psychiatry* 46, 266–280.

Thompson, B. (2004). *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications.* Washington, DC: American Psychological Association.

Van Roy, B., Veenstra, M., and Clench-Aas, J. (2008). Construct validity of the five-factor Strengths and Difficulties Questionnaire (SDQ) in pre-, early, and late adolescence. *J. Child Psychol. Psychiatry* 49, 1304–1312.

van Widenfelt, B. M., Goedhart, A. W., Treffers, P. D. A., and Goodman, R. (2003). Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *Eur. Child. Adolesc. Psychiatry* 12, 281–289.

Verhulst, F. C., van der Ende, J., Ferdinand, R. F., and Kasius, M. C. (1997). The prevalence of DSM-III-R diagnoses in a national sample of Dutch adolescents. *Arch. Gen. Psychiatry* 54, 329–336.

Vostanis, P. (2006). Strengths and Difficulties Questionnaire: research and clinical applications. *Curr. Opin. Psychiatry* 19, 367–372.

Woerner, W., Becker, A., and Rothenberger, A. (2004a). Normative data and scale properties of the German parent SDQ. *Eur. Child. Adolesc. Psychiatry* 13, II3–II10.

Woerner, W., Fleitlich-Bilyk, B., Martinussen, R., Fletcher, J., Cucchiaro, G., Dalgalarrondo, P., Lui, M., and Tannock, R. (2004b). The Strengths and Difficulties Questionnaire overseas: evaluations and applications of the SDQ beyond Europe. *Eur. Child. Adolesc. Psychiatry* 13(Suppl. 2), II47–II54.

Yasong, D., Jianhua, K., and Coghill, D. (2008). The validity, reliability and normative scores of the parent, teacher and self report versions of the Strengths and Difficulties Questionnaire in China. *Child Adolesc. Psychiatry Ment. Health* 2, 8.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.