Journal of Adolescence 38 (2015) 49–56

Contents lists available at ScienceDirect

Journal of Adolescence

journal homepage: www.elsevier.com/locate/jado

Screening mental health problems during adolescence: Psychometric properties of the Spanish version of the Strengths and Difficulties Questionnaire

Javier Ortuño-Sierra [a], Eduardo Fonseca-Pedrero [a, *], Mercedes Paino [b], Sylvia Sastre i Riba [a], José Muñiz [b, c]

[a] Department of Educational Sciences, University of La Rioja, Spain
[b] Department of Psychology, University of Oviedo, Spain
[c] Center for Biomedical Research in the Mental Health Network (CIBERSAM), Spain

Keywords:
SDQ
Self-report
Adolescents
Factorial structure
Measurement invariance
Behavioural problems

A B S T R A C T

The main purpose of the present study was to test the psychometric properties of the Strength and Difficulties Questionnaire (SDQ), self-reported version, in Spanish adolescents, introducing a five-point Likert response scale. The sample consisted of 1474 adolescents with a mean age of 15.92 years (*SD* = 1.18). The level of internal consistency of the SDQ Total score was .75, ranging from .56 to .71 for the subscales. Results from exploratory factor analysis revealed a three-factor structure as the most satisfactory. Confirmatory factor analyses showed that the five-factor model (with modifications) displayed better goodness of-fit indices than the other hypothetical dimensional models tested. Furthermore, strong measurement invariance by age and partial measurement invariance by gender was supported. The study of the psychometric properties confirms that the Spanish version of the SDQ, self-reported form, is a useful tool for the screening of emotional and behavioural problems in adolescents.

Introduction

Mental health problems in children and adolescents have an important impact not only in the individual but also in the family, in the school environment, and in the public global health (Gore et al., 2011; Meltzer, Gatward, Goodman, & Ford, 2003). Interest in the detection of children and adolescents at risk for emotional disorders or behavioural problems has increased in the last two decades (Carli et al., 2014; Erol, Simsek, Oner, & Munir, 2005; Kessler et al., 2012; Merikangas et al., 2010). Despite the efforts in early detection, different studies have suggested that only a minority of the adolescent population with needs in the area of mental health comes in direct contact with specialized services (Angold et al., 1998; Ford, Hamilton, Meltzer, & Goodman, 2008).

The assessment of emotional and behavioural problems in children and adolescents is a priority issue not only for public health policy, but also in the context of clinical practice and research. Standardized assessment by means of self-report allows

* Corresponding author. Department of Educational Sciences, University of La Rioja, C/ Luis de Ulloa, s/n, Edificio VIVES, PC: 26002, Logroño, La Rioja, Spain. Tel.: +34 941 299 309; fax: +34 941 299 333.
E-mail address: eduardo.fonseca@unirioja.es (E. Fonseca-Pedrero).

http://dx.doi.org/10.1016/j.adolescence.2014.11.001
0140-1971/

the exploration of prevalence rates, the frequency and the distribution of psychological symptoms and disorders, and the testing of the underlying structure of empirically based taxonomies (Fonseca-Pedrero, Sierra-Baigrie, Lemos Giraldez, Paino, & Muñiz, 2012). The Strength and Difficulties Questionnaire (SDQ) (Goodman, 1997) is a screening instrument for behavioural and emotional problems that similarly allows the assessment of capacities in the social sphere. Furthermore, it is a brief, simple, and easy management tool for use in child and adolescent populations (Ruchkin, Jones, Vermeiren, & Schwb-Stone, 2008; Vostanis, 2006). The SDQ is composed of 25 items in a Likert response format with three options grouped into five subscales or dimensions (Goodman, 1997): Emotional symptoms, Conduct problems, Hyperactivity, Peer problems, and Prosocial behaviour. The first four subscales form a Total difficulties score. The items that compose the SDQ are both positively and negatively phrased in order to avoid the effect of response bias (e.g., acquiescence). In total, 15 items reflect problems and 10 capabilities, of which five belong to the Prosocial subscale and five should be recoded, since they belong to the difficulties subscale.

Previous studies have reported adequate psychometric properties related to reliability and sources of validity evidences for the SDQ self-reported version (Gómez, 2012; Klasen et al., 2000; Muris, Meesters, & van den Berg, 2003). Nevertheless, several studies have detected low values of reliability through Cronbachs's alpha coefficient ($\alpha < .60$), especially in the subscales of Conduct problems and Peer problems (Becker, Hagenberg, Roessner, Woerner, & Rothenberg, 2004; Capron, Therond, & Duyme, 2007; Goodman, 2001; Koskelainen, Sourander, & Kaljonen, 2000; Mellor, 2004; Mellor & Stokes, 2007; Muris & Maas, 2004; Rønning, Helge Handegaard, Sourander, & Mørch, 2004; Ruchkin, Koposov, & Schwab-Stone, 2007; Yao et al., 2009). It is also noteworthy that the original format's response of the SDQ is a Likert type with three response options, which may also contribute to the low levels of reliability found. Previous studies have indicated that Likert response format, with few response options, have lower levels of reliability (Zumbo, Gadermann, & Zeisser, 2007). Moreover, the reformulation of the items in positive terms, could be a key factor in explaining low levels in Cronbach's alpha coefficient and the inconsistency of factorial solutions (van de Looij-Jansen, Goedhart, de Wilde, & Treffers, 2011). The fact that the problems subscales includes this type of items can generate that they behave as part of a distinct construct (Goodman, 2001). Therefore, reverse-worded items may influence the estimation of internal consistency due to their low correlation with the rest of the SDQ items that measure problems, and could, at the same time, affect the factor structure (van de Looij-Jansen et al., 2011).

Studies of the factor structure of the SDQ self-reported version yielded contradictory results. Previous studies, conducted using exploratory factor analysis, found support for the original five-factor structure (Goodman, 2001; Koskelainen, Sourander, & Vauras, 2001; Muris, Meesters, Eijkelenboom, & Vincken, 2004), while others reported a three-factor structure (Koskelainen et al., 2001; Percy, McCrystal, & Higgins, 2008; Ruchkin et al., 2008), and even a four-factor solution being most satisfactory (Muris et al., 2004). Regarding the confirmatory factor analysis, previous studies showed the five-factor solution as the most appropriate (He, Burstein, & Schmitz, 2013; Ruchkin et al., 2008; Ruchkin et al., 2007; Svedin & Priebe, 2008; Van Roy, Veenstra, & Clench-Aas, 2008; Yao et al., 2009), while others found the three-factor solution (Percy et al., 2008; Ruchkin et al., 2008), or even a five-factor solution with two second order factor (internalizing and externalizing) as the most satisfactory (Goodman, Lamping, & Ploubidis, 2010). Nevertheless, Mellor and Stokes (2007) reported that none of the five subscales was essentially one-dimensional, questioning the adequacy of the internal structure of the five-factor solution. Other research, likewise, discussed the adequacy of the setting of subscales, concluding that the SDQ factorial structure was not appropriate (Percy et al., 2008; Rønning et al., 2004).

Another important issue regarding the factor structure of the SDQ is the study of measurement invariance across groups. The evaluation of measurement invariance is important for determining the generalizability of latent constructs across groups and whether the measurement instrument and the construct being measured are operating in the same way across diverse samples of interest (Byrne, 2012). If measurement invariance does not hold, inferences and interpretations drawn from the data may be erroneous or unfounded. Different studies have analysed the measurement invariance of the SDQ, self-reported version in adolescents, across different variables (e.g., gender, age, race/ethnicity, and income) (He et al., 2013; Rønning et al., 2004; Ruchkin et al., 2008; van de Looij-Jansen et al., 2011). As yet, there has been no in-depth examination of the question of whether or not the dimensional structure of the SDQ is invariant across gender and age.

Although a significant number of investigations have studied the psychometric properties of the SDQ in Europe as well as in America and Asia, there are few studies in the review of the literature that analyse the psychometric quality of the SDQ in its Spanish version. Therefore, the main purpose of the present study was to study the psychometric properties of the SDQ scores, self-reported version, with a five Likert response format in a representative sample of Spanish adolescents. From this general goal three specific objectives have been formulated: a) to examine the internal consistency of the SDQ scores through Cronbach's alpha; b) to study the dimensional structure of the SDQ scores using exploratory and confirmatory factor analyses; and c) to test the measurement invariance of the SDQ scores across gender and age. Based in previous research and results, it is hypothesised that sound reliability will be established, and that the proposed five factor dimensional model will be supported. It is further hypothesised that the five factor dimensional model of SDQ will be equivalent by age and gender.

Method

Participants

Selection of participants was by means of stratified random sampling by clusters, at the classroom level, in a population of approximately 36,000 students from the Principality of Asturias (a region situated in the north of Spain). Strata were created according to geographical area – East, West, Central and South – and educational level – compulsory and post-compulsory –, as it applies to Spanish educational system and the probability of a school being selected depended on the number of students. Pupils were from different types of secondary schools – public, grant-assisted private, and private – and from vocational/technical schools. The initial sampling was formed by 1628 students, eliminating participants who presented: a) a high score in The Oviedo Infrequency Scale (more than two points) ($n = 64$); b) learning difficulties ($n = 6$); c) omission of demographics or a high percentage of items without responding ($n = 48$); and d) outlier scores ($n = 36$). Thus, the final sample comprised a total of 1474, 756 were male students (48.6%), belonging to 28 schools and 90 classrooms. The age of the participants ranged from 14 to 18 ($M = 15.92$; $SD = 1.18$). The age distribution of the sample was the following: 14 years ($n = 194$; 14.7%), 15 years ($n = 357$; 27.1%), 16 years ($n = 411$; 31.2%), 17 years ($n = 357$; 27.1%) and 18 years ($n = 137$; 9.3%). With the aim of conducting pertinent statistical analyses, a cross-validation study was performed where the total sample was randomly split into two subsamples. The first sub-sample consisted of 765 participants (379 male), with a mean age of 15.87 years ($SD = 1.18$). The second sub-sample consisted of 709 participants (337 male), mean age of 15.98 years ($SD = 1.17$). Neither gender ($\chi^2 = 0.106$; $p = .744$) nor age rates ($F = 0.286$; $p = .836$) differed across subsamples.

Instruments

The Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997), self-reported form. It is a measuring instrument widely used for the assessment of different social, emotional, and behavioural problems related to mental health in children and adolescents over the previous 6 months. The SDQ is made up of a total of 25 statements distributed across five subscales (each with five items): Emotional symptoms, Conduct problems, Hyperactivity, Peer problems, and Prosocial behaviour. In this study we used a Likert-type response format with five options (1 = "*totally disagree*" to 5 = "*totally agree*"), so that the score on each subscale ranged from 5 to 25 points. In the present study we used the version adapted and translated into Spanish in a non-clinical adolescent population (Fonseca-Pedrero, Paino, Lemos-Girádez, & Muñiz, 2011; Ortuño-Sierra, Fonseca-Pedrero, Paíno, & Aritio-Solana, 2014).

The Oviedo Infrequency Scale (INF-OV) (Fonseca-Pedrero, Lemos-Giráldez, Paino-Pineiro, Villazón-García, & Muñiz, 2009). INF-OV is a 12-item self-report instrument with a Likert-type response format using five categories (1 = "*totally disagree*" to 5 = "*totally agree*"). Its objective is to detect those participants who respond to self-reports in a random, pseudo-random or dishonest fashion. Once the items were dichotomized, participants who scored more than two items incorrectly were eliminated from the study.

Procedure

The questionnaires were administered collectively, in groups of 10–35 students, during normal school hours and in a classroom specially prepared for this purpose. For participants under 18, parents were asked to provide written informed consent in order for their child to participate in the study. Participants were informed of the confidentiality of their responses and of the voluntary nature of the study. No incentive was provided for their participation. Administration took place under the supervision of researchers. The study was approved by the research and ethics committees at the University of Oviedo and by the Department of Education of the Principality of Asturias.

Data analyses

First, we calculated descriptive statistics (mean, standard deviation, skewness, and kurtosis). Second, we examined internal consistency of the SDQ subscales and Total score. Third, in order to analyse the internal structure of SDQ scores, we conducted a cross-validation study, randomly dividing the total sample into two subsamples. In the first subsample, exploratory factor analyses were performed using the Unweighted Least Squares. The Pearson correlation matrix was used. In the second subsample, several confirmatory factor analyses (CFAs) were conducted. The parameters were obtained from the Muthen's quasi-likelihood estimator (Muthén & Muthén, 1998–2007). The following goodness-of-fit indices were used: Chi-square (χ^2), Confirmatory Factor Index (CFI), Tucker–Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). Hu and Bentler (1999) suggested that RMSEA should be .06 or less for a good model fit and CFI and TLI should be .95 or more, though any value over .90 tends to be considered acceptable. For SRMR, values less than .08 indicate good model fit.

Finally, in order to test measurement invariance (MI), successive multigroup CFAs were conducted (Byrne, 2008). Basically, a hierarchical set of steps are followed when MI is tested, typically starting with the determination of a well-fitting multi-group baseline model and continuing with the establishment of successive equivalence constraints in the model parameters across groups. The basal model is called the configural model, which is the first and least restrictive model to be tested. The configural model is established by specifying and testing the model for each group separately. Once the theoretical model has been validated in both groups, configural invariance is examined requiring that the same pattern of fixed and freely estimated parameters are equivalent across groups, and therefore, that no equality constraints are imposed. Metric or weak invariance is tested, where the equivalence of the factorial loadings across groups is tested. Factor loadings are freely estimated for the first group only, and in the remaining groups these parameters estimated are constrained equal to those of the first group. Finally, strong or scalar MI is tested, where the item intercepts and the factor loadings are equally constrained across groups. The analysed dimensional models can be seen as nested models to which constraints are progressively added. Due to the limitations of the $\Delta\chi^2$ regarding its sensitivity to sample size, Cheung and Rensvold (2002) proposed a more practical criterion, the change in CFI (ΔCFI), to determine if nested models are practically equivalent. In this study, when ΔCFI is greater than .01 between two nested models, the more constrained model is rejected since the additional constraints have produced a practically worse fit. However, if the change in CFI is less than or equal to .01, it is considered that all specified equal constraints are tenable, and therefore, it is possible to continue with the next step in the analysis of MI. SPSS 15.0 (Statistical Package for the Social Sciences, 2006), Mplus 5.0 (Muthén & Muthén, 1998–2007), and FACTOR 9.2 (Lorenzo-Seva & Ferrando, 2006) were used for data analysis.

Results

Descriptive statistics for the SDQ subscales

Descriptive statistics for the SDQ subscales and Total score for the total sample and by gender are shown in Table 1. As shown in Table 1, the internal consistency of the Total difficulties score was .75 for the whole sample (.76 male and .75 female).

Validity evidence based on internal structure of the SDQ scores: exploratory and confirmatory factor analysis

Exploratory factor analyses were conducted using the first subsample. The KMO measure of sampling adequacy was .78, and the Bartlett test of sphericity was 3481.8 ($p < .001$). The results suggested a three-dimension factor solution as the most parsimonious. The selection of three factors was selected as the most accurate in terms of psychological interpretation, the scree plot, and the parsimony criterion. Moreover, the analysis of the factor loadings suggested the following three factors: Internalizing problems (15.91% explained variance), Externalizing (10.13% explained variance) problems, and Prosocial (9.03% explained variance). For this three-factor structure, the goodness-of-fit indices were GFI = .97 and AGFI = .96. As shown in Table 2, the item distribution is not entirely homogeneous and some overlaps were found between factors. Items 12, 18, and 22 of the Conduct problems subscale showed a higher factor loading in the Prosocial subscale, as well as item 11 pertaining to the Peer problems subscale. The correlations between factors ranged from .20 (FI-FIII) to −.08 (FI-FII) ($p < .01$).

Different hypothetical dimensional models were tested in the confirmatory factor analysis: a) the three-factor model suggested in the EFA (with and without modification); b) the five-factor model (with and without modification) (Goodman, 1997); c) two models with prosocial factor influence extended from the original models of three and five factors (van de Looij-Jansen et al., 2011); and d) the five-factor model with 2 s-order factors (Goodman et al., 2010), resulting from grouping internalizing symptoms (emotional and peer) and externalizing symptoms (behavioural and hyperactive). As shown in Table 3, goodness-of-fit indices for the three-factor baseline model did not reach the cut-offs recommended. The five-factor baseline model showed better goodness-of-fit index but was still questionable. For both models, substantial Modification Indices (MIs) were found, for error correlation between items 25 and 15, items 2 and 10, items 19 and 18, and items 15 and 16. This correlation between error terms was made between those items that have similar content. Confirmatory factor analyses showed that the five-factor model (with modifications) displayed better goodness of-fit indices than the other hypothetical dimensional models tested. Meanwhile, the model with the inclusion of second-order factors revealed lower goodness-of-fit indices than the five-factor model. The standardized factor loadings for the five-factor model allowing correlation between the error terms are shown in Table 4.

Measurement invariance of the SDQ scores across gender and age

To examine measurement invariance across age, the sample was divided into two subgroups (14–16 year-olds and 16–18 year-olds), according to the stages of the Spanish educational system (compulsory/post-compulsory). Prior to the analysis of measurement invariance across gender and age, we tested whether the five-factor model with modifications showed a reasonable good fit in each group. Next, configural, weak, and strong invariance across gender and age of participants was examined (see Table 5). Differences CFI below .01 between the configural model and the other models confirmed strong invariance for gender. In the case of age, ΔCFI above .01 were found; this leads to rejecting the hypothesis of strong invariance

Table 1
Descriptive statistics of the SDQ for gender and whole sample.

	Male (n = 716)				Female (n = 758)				Total (n = 1474)			
	M (*SD*)	Skewness	Kurtosis	Alpha	*M* (*SD*)	Skewness	Kurtosis	Alpha	*M* (*SD*)	Skewness	Kurtosis	Alpha
Emotional	11.11 (3.52)	0.70	0.44	.65	13.15 (4.11)	0.23	−0.35	.71	12.16 (3.96)	0.48	−0.15	.71
Conduct	11.12 (3.33)	0.78	0.95	.58	9.65 (2.80)	0.67	0.47	.53	10.36 (3.15)	0.80	1.00	.58
Peer	9.33 (2.98)	1.27	2.43	.57	9.21 (2.83)	1.12	1.76	.55	9.26 (2.90)	1.21	2.13	.56
Hiperactivity	14.59 (3.99)	0.15	−0.27	.70	14.14 (3.67)	0.13	−0.07	.65	14.36 (3.83)	0.15	−0.16	.68
Prosocial	19.76 (2.79)	−0.65	1.22	.64	20.90 (2.54)	−0.59	0.23	.62	20.34 (2.72)	−0.64	0.84	.64
Total score	46.15 (9.31)	0.55	1.09	.77	46.15 (8.92)	0.26	0.03	.76	46.15 (9.11)	0.41	0.59	.75

Table 2
Exploratory factor analysis of the SDQ items.

SDQ items	Factors		
	I	II	III
3		.36	
8		.57	
13		.66	
16		.52	
24		.56	
5			.40
7	−.33		.39
12			.21
18			
22			.23
6		.39	
11	−.30	.26	
14		.40	
19		.48	
23		.36	
2			.63
10			.67
15			.53
21			.43
25			.40
1	.60		
4	.45		
9	.58		
17	.45		
20	.51		

Note. Factorial loadings under .30 have been omitted.
Factorial loadings between .20 and .29 in shaded letters.

for this variable. Based on Cheung and Rensvold (2002), we conducted a series of successive analysis to locate the intercepts of the items causing ΔCFI (1, 3, 4, 5, 9, 13, 24, 14, 15, 16, 18, 22, and 24). Once the item parameters were released, partial measurement invariance was supported for gender.

Discussion and conclusions

The main purpose of this study was to analyse the psychometric quality of the Strength and Difficulties Questionnaire (SDQ) (Goodman, 1997) in its self-reported form with a five point Likert response format in a representative sample of Spanish adolescents. To this end, we estimated the reliability of the SDQ scores, examined the internal structure, and tested the measurement invariance by gender and age. Knowledge of the SDQ psychometric properties is relevant for use it as a screening tool in an age group at particular risk of developing emotional and behavioural symptoms and disorders (Carli et al., 2014; Erol et al., 2005; Kessler et al., 2012; Merikangas et al., 2010).

The SDQ scores showed discrete reliability levels in Conduct and Peer problems subscales. Adequate levels of reliability were found with Cronbach's alpha for the Total score of .75. Cronbach's alpha for female and male in the Total difficulties score ranged from .53 to .77. As is the case in previous studies (Becker et al., 2004; Goodman, 2001; Koskelainen et al., 2000; Mellor & Stokes, 2007; Muris et al., 2004; Rønning et al., 2004; Ruchkin et al., 2008; Ruchkin et al., 2007; Yao et al., 2009), Conduct and Peer problems subscales showed lower internal consistency levels, with values of .58 and .56 for the total sample.

Table 3
Goodness-of-fit indices of the models tested in the confirmatory factor analysis.

Models	χ^2	*df*	CFI	TLI	RMSEA	SRMR
Baseline 5-factor	1685.101	265	.831	.808	.046	.042
Baseline 5-factor with reverse-worded items added to prosocial factor	1502.612	260	.852	.829	.043	.038
Final 5-factor with correlated errors added: 19–18, 2–10, 25–15, 15–16	1103.285	261	.900	.885	.036	.036
Final 5-factor with reverse-worded items added to prosocial factor: 19–18, 2–10, 25–15, 15–16	966.710	256	.915	.901	.033	.033
Baseline 3-factor	2541.061	272	.730	.702	.057	.054
Baseline 3-factor with reverse-worded items added to prosocial factor	2300.472	267	.758	.728	.055	.050
Final 3-factor with correlated errors added: 19–18, 2–10, 25–15, 15–16	1807.676	268	.816	.795	.047	.048
Final 3-factor with correlated errors added and reverse-worded items added to prosocial factor: 19–18, 2–10, 25–15, 15–16	1582.566	263	.843	.821	.044	.044
Goodman et al., (2010) 5-factor and 2 second-order factor	1800.492	268	.817	.796	.047	.046

Note. χ^2 = Chi square; *df* = degrees of freedom; CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR= Standardized Root Mean Square Residual.

Table 4
Standardized factor loadings for the five-factor final model.

Items	Loadings	R^2
Emotional problems		
3	.43	.19
8	.60	.37
13	.72	.53
16	.55	.31
24	.54	.29
Conduct problems		
5	.54	.29
7	.47	.22
12	.51	.26
18	.44	.19
22	.43	.19
Peer problems		
6	.48	.23
11	.40	.16
14	.57	.33
19	.48	.23
23	.33	.11
Hiperactivity		
2	.45	.20
10	.47	.22
15	.54	.30
21	.51	.26
25	.41	.17
Prosocial		
1	.64	.41
4	.45	.21
9	.54	.29
17	.44	.20
20	.53	.29

Note. All standardized factorial loadings estimated were statistically significant ($p < .01$).

Cronbach's alpha values were acceptable but in some cases did not reach appropriate values recommended above .70 (Nunnally, 1978). One reason to explain this could be in the homogeneity of the sample used in our study. In addition, the fact that each SDQ subscale is composed by only five items is another variable that may affect the reliability. Moreover, the low levels of reliability found in some subscales as it is the case of the Peer problems may be affecting the total score reliability. As has been proposed, some items of this subscale seem to reflect loneliness (e.g., rather solitary, tends to play alone; has at least one good friend) and sociability (e.g., generally liked by other children; gets on better with adults than with other children) rather than peer problems (Stone, Otten, Engels, Vermulst, & Janssens, 2010). Furthermore, this particular subscale is more likely to be affected by social desirability, which may affect its reliability. It is noteworthy that possible improvement of the reliability of the SDQ scores could be determined by the five Likert response format used in this work. In the specialized literature, the use of this response format is recommended in order to improve the reliability of scores (Lozano, García-Cueto, & Muñiz, 2008; Muñiz, García-Cueto, & Lozano, 2005), as well as for dimensional scores on psychopathology measures (Markon, Chmielewski, & Miller, 2011).

Table 5
Goodness-of-fit indices for measurement invariance of the SDQ (five-factor model with modifications) across gender and age.

	χ^2	*df*	CFI	TLI	RMSEA	SRMR	ΔCFI
Gender							
Male ($n = 337$)	397.537	262	.895	.880	.039	.060	
Female ($n = 372$)	442.452	261	.870	.851	.043	.059	
Configural invariance	897.979	526	.861	.842	.045	.063	
Weak factorial invariance	949.773	551	.851	.838	.045	.070	−.01
Strong factorial invariance	965.617	563	.821	.840	.045	.071	+.01
Strong partial factorial invariance (freeing intercepts:1, 3, 4, 5, 9, 13, 24, 14, 15, 16, 18, 22, and 24)	965.617	563	.850	.840	.045	.071	−.01
Age Dichotomized (14–16 and 17–18 years)							
14–16 years ($n = 456$)	457.767	261	.883	.866	.041	.058	
17–18 years ($n = 253$)	365.767	261	.898	.882	.041	.064	
Configural invariance	821.971	522	.889	.872	.040	.060	
Weak factorial invariance	840.876	547	.891	.881	.059	.063	−.01
Strong factorial invariance	882.775	572	.885	.879	.039	.064	−.01

Note. χ^2 = Chi square; *df* = degrees of freedom; CFI = Confirmatory Factor Index; TLI = Tucker–Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; ΔCFI = Change in Confirmatory Factor Index.

In the study of the internal structure of SDQ, successive exploratory factor analyses revealed a three-factor structure to be the most appropriate. However, confirmatory factor analyses (CFAs) supported the five-factor structure, as it is the case in previous studies (He et al., 2013; Ruchkin et al., 2008; Ruchkin et al., 2007; Svedin & Priebe, 2008; Van Roy et al., 2008; Yao et al., 2009). Nevertheless, optimal levels of goodness-of-fit indices were found after adding error correlation between items, indicating discrete values in the five-factor baseline model. Similar results were found in previous studies (Percy et al., 2008; Rønning et al., 2004). The results of CFAs rejecting the proposed three-factor model was found as well in different studies (Koskelainen et al., 2001; Percy et al., 2008; Ruchkin et al., 2008). Regarding this, ΔCFI analysis revealed that, unlike what has been reported by van de Looij-Jansen et al., (2011), both in the three and the five-factor models, the MIs correction is more significant in model fit than the inclusion of the extended Prosocial subscale with reverse-worded items. Taking everything into consideration, the original five factor structure is recommended as it displayed better goodness-of-fit indices and allows determining more psychological difficulties. In addition, the Emotional symptoms and Peer problems subscales have showed in previous studies (van de Looij-Jansen et al., 2011) different associations with gender, educational level, and ethnicity, which could reflect that these subscales represent substantially different and relevant areas that have to be considered. Nonetheless, in all cases, the extent of the Prosocial subscale resulted in an improvement of the model fit for both models, confirming the results of van de Looij-Jansen et al., (2011), and the idea that the extended prosocial factor may reflect the possibility of a positive response construct (Goodman, 2001). Furthermore, analysis of the MIs from the CFAs, suggested the existence of a minor factor in the Hyperactivity subscale.

Results supported the hypothesis of strong measurement invariance by age and partial measurement invariance by gender. The lack of strong invariance by gender might indicate differential item functioning in this variable. The finding of measurement equivalence across age and gender provides essential evidence of construct validity for the SDQ scores. Adolescence is a developmental stage in which relevant biopsychological changes occur (Lerner & Galambos, 1998). These changes are different for male and female and do not happen at the same time. For this reason, we believe that the study of the measurement invariance is relevant in order to assure the comparability of scores and for determining the generalizability of latent constructs across groups. The review of the literature shows that there are few studies of measurement invariance in the self-reported version of the SDQ (He et al., 2013; Rønning et al., 2004; Ruchkin et al., 2008; van de Looij-Jansen et al., 2011). Recent studies have found partial measurement invariance in the SDQ self-reported version in adolescents, across different demographic variables including gender and age. For instance, van de Looij-Jansen et al., (2011) showed that self-reported version of the SDQ was invariant by age, education level, and ethnicity, while the hypothesis of strong factorial invariance across gender was not clearly acceptable. Rønning et al., (2004) confirmed invariance across gender for the SDQ scores, although the initial setting of the model in men and women was inappropriate, while He et al., (2013) found invariance across gender, age, race/ethnicity, and income subgroups.

The results of the present study should be interpreted in the light of the following limitations. One possible limitation of this study is that in spite of having a representative sample of adolescents, we focused on a particular Spanish region. Given the peculiarities, diversity and plurality of the nation, future studies could examine the psychometric properties of the instrument in an adolescent population in other regions or geographic areas. Future studies could replicate the study of the psychometric properties of the SDQ in its Spanish version. Also, future research on the measurement invariance across cultures, in the self-reported version, would allow the comparison of results between different countries and/or cultures.

Acknowledgements

This research was funded by the Spanish Ministry of Science and Innovation (MICINN) and by the Instituto Carlos III, Center for Biomedical Research in the Mental Health Network (CIBERSAM). Project references: PSI 2011-28638, PSI 2011-23818.

References

Angold, A., Messer, S. C., Stangl, D., Farmer, E., Costello, E. J., & Burns, B. J. (1998). Perceived parental burden and service use for child and adolescent psychiatric disorder. *American Journal of Public Health, 88*, 75–80.

Becker, A., Hagenberg, N., Roessner, N., Woerner, W., & Rothenberg, A. (2004). Evaluation of the self-reported SDQ in a clinical setting: do self-reports tell us more than ratings by adult informants? *European Child & Adolescent Psychiatry, 13*(2), 17–24. http://dx.doi.org/10.1007/s00787-004-2004-4.

Byrne, B. (2008). Testing for multigroup equivalence of a measuring instrument: a walk through the process. *Psicothema, 20*, 872–882.

Byrne, B. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York: Routledge Taylor & Francis Group.

Capron, C., Therond, C., & Duyme, M. (2007). Psychometric properties of the French version of the self-report and teacher Strengths and Difficulties Questionnaire (SDQ). *European Journal of Psychological Assessment, 23*, 79–88. http://dx.doi.org/10.1027/1015-5759.23.2.79.

Carli, V., Hoven, C. W., Wasserman, C., Chiesa, F., Guffanti, G., Sarchiapone, M., et al. (2014). A newly identified group of adolescents at "invisible" risk for psychopathology and suicidal behavior: findings from the SEYLE study. *World Psychiatry, 13*, 78–86.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233–255. http://dx.doi.org/10.1207/S15328007SEM0902_5.

Erol, N., Simsek, Z., Oner, O., & Munir, K. (2005). Behavioral and emotional problems among Turkish children at ages 2 to 3 years. *Journal of the American Academic of Child and Adolescence Psychiatry, 44*, 80–87. http://dx.doi.org/10.1097/01.chi.0000145234.18056.82.

Fonseca-Pedrero, E., Lemos-Giráldez, S., Paino-Pineiro, M., Villazón-García, U., & Muñiz, J. (2009). Validation of the Schizotypal Personality Questionnaire-brief form in adolescents. *Schizophrenia Research, 111*, 53–60.

Fonseca-Pedrero, E., Paino, M., Lemos-Girádez, S., & Muñiz, J. (2011). Prevalencia de la sintomatología emocional y comportamental en adolescentes españoles a través del Strengths and Difficulties Questionnaire (SDQ) [Prevalence of emotional and behavioral symptoms in Spanish adolescents through the Strengths and Difficulties Questionnaire (SDQ)]. *Revista de Psicopatología y Psicología Clínica, 16*, 15–25.

Fonseca-Pedrero, E., Sierra-Baigrie, S., Lemos Giraldez, S., Paino, M., & Muñiz, J. (2012). Dimensional structure and measurement invariance of the Youth Self-Report across gender and age. *Journal of Adolescent Health, 50*, 148–153.

Ford, T., Hamilton, H., Meltzer, H., & Goodman, R. (2008). Predictors of service use of mental health problems among British school children. *Child and Adolescent Mental Health, 13*, 32–40.

Gómez, R. (2012). Correlated trait-correlated method minus one analysis of the convergent and discriminant validities of the Strengths and Difficulties Questionnaire. *Assesment, 12*, 372–382.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry, 38*, 581–586.

Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academic of Child and Adolescence Psychiatry, 40*, 1337–1345.

Goodman, A., Lamping, D. L., & Ploubidis, G. B. (2010). When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): data from British parents, teachers and children. *Journal of Abnormal Child Psychology, 38*, 1179–1191.

Gore, F. M., Bloem, P. J., Patton, G. C., Ferguson, J., Joseph, V., Coffey, C., et al. (2011). Global burden of disease in young people aged 10–24 years: a systematic analysis. *Lancet, 18*, 2093–2102.

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

Hu, L.-T., & Bentler, P. M. (1999). Cut off criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Kessler, R. C., Avenevoli, S., Costello, E. J., Georgiades, K., Green, J. G., Gruber, M. J., et al. (2012). Prevalence, persistence, and sociodemographic correlates of DSM-IV disorders in the National Comorbidity Survey Replication Adolescent Supplement. *Archives of General Psychiatry, 69*, 372–380.

Klasen, H., Woerner, W., Wolke, D., Meyer, R., Overmeyer, S., Kaschnitz, W., et al. (2000). Comparing the German versions of the Strengths and Difficulties Questionnaire (SDQ-Deu) and the child behavior checklist. *European Child and Adolescent Psychiatry, 9*, 271–276.

Koskelainen, M., Sourander, A., & Kaljonen, A. (2000). The Strenght and Difficulties Questionnaire among finnish school-aged children and adolescents. *European Child and Adolescent Psychiatry, 9*, 277–284.

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

Lerner, R. M., & Galambos, N. L. (1998). Adolescent development: challenges and opportunities for research, programs, and policies. *Annual Review of Psychology, 49*, 413–446.

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

Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: a computer program to fit the exploratory factor analysis model. *Behavior Research Methods Instruments & Computers, 38*, 88–91.

Lozano, L. M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology, 4*, 73–79.

Markon, K. E., Chmielewski, M., & Miller, C. J. (2011). The reliability and validity of discrete and continuous measures of psychopathology: a quantitative review. *Psychological Bulletin, 137*, 856–879.

Mellor, D. (2004). Furthering the use of the Strengths and Difficulties Questionnaire: reliability with younger child respondents. *Psychological Assessment, 16*(4), 396–401. http://dx.doi.org/10.1037/1040-3590.16.4.396.

Mellor, D., & Stokes, M. (2007). The factor structure of the Strengths and Difficulties Questionnaire. *European Journal of Psychological Assessment, 23*(2), 105–112. http://dx.doi.org/10.1027/1015-5759.23.2.105.

Meltzer, H., Gatward, R., Goodman, R., & Ford, T. (2003). Mental health of children and adolescents in Great Britain. *International Review of Psychiatry, 15*(1–2), 185–187.

Merikangas, K. R., He, J. P., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., et al. (2010). Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry, 49*, 980–989.

Muñiz, J., García-Cueto, E., & Lozano, L. M. (2005). Item format and the psychometric properties of the Eysenck Personality Questionnaire. *Personality and Individual Differences, 38*, 61–69.

Muris, P., & Maas, A. (2004). Strengths and difficulties as correlates of attachment style in institutionalized and non-institutionalized children with below-average intellectual abilities. *Child Psychiatry and Human Development, 34*, 317–328.

Muris, P., Meesters, C., & van den Berg, F. (2003). The Strengths and Difficulties Questionnaire (SDQ). Further evidence for its reliability and validity in a community sample of Dutch children and adolescents. *European Child & Adolescent Psychiatry, 12*, 1–8. http://dx.doi.org/10.1007/s00787-003-0298-2.

Muris, P., Meesters, C., Eijkelenboom, A., & Vincken, M. (2004). The self-report version of the Strengths and Difficulties Questionnaire: its psychometric properties in 8- to 13- year-old non-clinical children. *British Journal of Clinical Psychology, 43*(Pt4), 437–448. http://dx.doi.org/10.1348/0144665042388982.

Muthén, L. K., & Muthén, B. O. (1998-2007). *Mplus user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Ortuño-Sierra, J., Fonseca-Pedrero, E., Paíno, M., & Aritio-Solana, R. (2014). Prevalencia de síntomas emocionales y comportamentales en adolescentes españoles [Prevalence of emotional and behavioral symptomatology in Spanish adolescents]. *Revista de Psiquiatría y Salud Mental, 7*, 121–130. http://dx.doi.org/10.1016/j.rpsm.2013.12.003.

Percy, A., McCrystal, P., & Higgins, K. (2008). Confirmatory factor analysis of the adolescent self-report Strengths and Difficulties Questionnaire. *European Journal of Psychological Assessment, 24*, 43–48. http://dx.doi.org/10.1027/1015-5759.24.1.43.

Rønning, J. A., Helge Handegaard, B. H., Sourander, A., & Mørch, W.-T. (2004). The Strengths and Difficulties Self-Report Questionnaire as a screening instrument in Norwegian community samples. *European Child & Adolescent Psychiatry, 13*, 73–82. http://dx.doi.org/10.1007/s00787-004-0356-4.

Ruchkin, V., Jones, S., Vermeiren, R., & Schwb-Stone, M. (2008). The Strengths and Difficulties Questionnaire: the self-report version in American urban and suburban youth. *Psychological Assessment, 20*(2), 175–182.

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

Statistical Package for the Social Sciences. (2006). *SPSS Base 15.0 user's guide*. Chicago, IL: SPSS Inc.

Stone, L. L., Otten, R., Engels, R. C., Vermulst, A. A., & Janssens, J. M. (2010). Psychometric properties of the parent and teacher versions of the Strengths and Difficulties Questionnaire for 4- to 12-year-olds: a review. *Clinical Child and Family Psychology Review, 13*, 254–274.

Svedin, C. G., & Priebe, G. (2008). The Strengths and Difficulties Questionnaire as a screening instrument in a community sample of high school seniors in Sweden. *Nordic Journal of Psychiatry, 62*, 225–232. http://dx.doi.org/10.1080/08039480801984032.

Van Roy, B., Veenstra, M., & Clench-Aas, J. (2008). Construct validity of the five-factor Strengths and Difficulties Questionnaire (SDQ) in pre-, early, and late adolescence. *Journal of Child Psychology and Psychiatry, 49*(12), 1304–1312. http://dx.doi.org/10.1111/j.1469-7610.2008.01942.x.

Vostanis, P. (2006). Strengths and Difficulties Questionnaire: research and clinical applications. *Current Opinion in Psychiatry, 19*, 367–372. http://dx.doi.org/10.1097/01.yco.0000228755.72366.05.

Yao, S., Zhang, C., Zhu, X., Jing, X., McWhinnie, C. M., & Abela, J. R. Z. (2009). Measuring adolescent psychopathology: psychometric properties of the self-report Strengths and Difficulties Questionnaire in a sample of Chinese adolescents. *Journal of Adolescent Health, 45*, 55–62.

Zumbo, B. M., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods, 6*, 21–29.