

Does the Strengths and Difficulties Questionnaire – self report yield invariant measurements across different nations? Data from the International Child Mental Health Study Group

D. Stevanovic^{1*}, R. Urbán², O. Atilola³, P. Vostanis⁴, Y. P. Singh Balhara⁵, M. Avicenna⁶, H. Kandemir⁷, R. Knez⁸, T. Franic⁹ and P. Petrov¹⁰

¹ Clinic for Neurology and Psychiatry for Children and Youth, Belgrade, Serbia

² Institute of Psychology, Eötvös Loránd University, Budapest, Hungary

³ Department of Behavioral Medicine, Lagos State University College of Medicine Ikeja, Lagos, Nigeria

⁴ School of Psychology, Leicester University, Leicester, UK

⁵ Department of Psychiatry, All India Institute of Medical Sciences, Ansari Nagar, New Delhi, India

⁶ Faculty of Psychology, State Islamic University Syarif Hidayatullah, Jakarta, Indonesia

⁷ Department of Child and Adolescents psychiatry, Harran University, Sanliurfa, Turkey

⁸ Department of psychiatry, University Hospital Centre Rijeka, Rijeka, Croatia

⁹ Child and Adolescent Psychiatry, School of Medicine, University of Split, Split, Croatia

¹⁰ Department of Child and Adolescent Psychiatry, University Hospital St. Marina, Varna, Bulgaria

Aims. This study evaluated the measurement invariance of the strengths and difficulties questionnaire (SDQ) self-report among adolescents from seven different nations.

Methods. Data for 2367 adolescents, aged 13–18 years, from India, Indonesia, Nigeria, Serbia, Turkey, Bulgaria and Croatia were available for a series of factor analyses.

Results. The five-factor model including original SDQ scales emotional symptoms, conduct problems, hyperactivity–inattention problems, peer problems and prosocial behaviour generated inadequate fit degree in all countries. A bifactor model with three factors (i.e., externalising, internalising and prosocial) and one general problem factor yielded adequate degree of fit in India, Nigeria, Turkey and Croatia. The prosocial behaviour, emotional symptoms and conduct problems factor were found to be common for all nations. However, originally proposed items loaded saliently on other factors besides the proposed ones or only some of them corresponded to proposed factors in all seven countries.

Conclusions. Due to the lack of a common acceptable model across all countries, namely the same numbers of factors (i.e., dimensional invariance), it was not possible to perform the metric and scalar invariance test, what indicates that the SDQ self-report models tested lack appropriate measurement invariance across adolescents from these seven nations and it needs to be revised for cross-country comparisons.

Received 5 January 2014; Revised 21 February 2014; Accepted 22 February 2014; First published online 30 April 2014

Key words: Adolescent, culture, invariance, psychopathology.

Introduction

There is a long-standing view that all children around the world follow similar patterns of biological and cognitive development, although there are marked individual differences in developmental rates, temperament and adaptive success among them (Achenbach *et al.* 2008). Within this developmental

framework, culture strongly shapes the environments in which children develop, what consequently might lead to specificities in mental health expressions across different culture groups (Nikapota & Rutter, 2008). Over the past decades, variations in rates of disorders across cultural groups were observed, mostly due to the presence of culture-specific mental disorders, differences in the manifestation of disorders, and differences in risk factors across cultural/ethnic groups (Nikapota & Rutter, 2008).

Much of what we currently know about child's mental health internationally is based on the two main assessment systems – the Achenbach System of

* Address for correspondence: Dr Dejan Stevanovic, Clinic for Neurology and Psychiatry for Children and Youth, Dr Subotića Starijeg 6a, 11000 Beograd Belgrade, Serbia.
(Email: dejanstevanovic@eunet.rs)

Empirically Based Assessment (ASEBA) and Strengths and Difficulties Questionnaire (SDQ) (Achenbach *et al.* 2008). Both systems consider a dimensional approach to child and adolescent mental health assessment and both emphasise cross-cultural perspectives with self-, parent and teacher rating scales developed in various languages (Achenbach, 1991a, b, c; Goodman, 1997, 2001; Goodman *et al.* 2004; Achenbach & Rescorla, 2007).

Using the youth self-report (YSR) from the ASEBA or the SDQ self-report, it was observed that the prevalence rates of general psychopathology that adolescents report differ substantially across nations/countries (Achenbach *et al.* 2008). For example, considering the SDQ only (Goodman, 1997, 2001), the rates of self-reported mental health problems in adolescent samples were 6.6% in Germany (Ravens-Sieberer *et al.* 2008a), 8.7% in Ireland (Greally *et al.* 2010), or 5.3% in the Gaza Strip (Thabet *et al.* 2000). However, in cross-cultural studies comparing analogous samples, a 1.6–2.8-fold difference in the rates had been observed across several countries (Ravens-Sieberer *et al.* 2008b; Lai *et al.* 2010; Atilola *et al.* 2013). The reasons why the prevalence rates estimated by self-reports differ substantially across various nations can be many. There might be inherent cross-cultural differences due to many different economic, social and cultural factors that contribute to the development and expression of specific psychopathology (e.g. Hackett & Hackett, 1999; Mabe & Josephson, 2004; Camras & Fatani, 2006; Nikapota & Rutter, 2008). There might be cross-cultural differences due to completion rates, recruitment methods or adolescent's age and development at assessments (Achenbach *et al.* 2012). Additionally, there might be cross-cultural differences in the SDQ or ASEBA measurement model itself, non-availability of population-specific norms, or inconsistencies in determining levels of psychopathology between a self-report questionnaire and clinical interview (Heiervang, *et al.* 2008; Achenbach *et al.* 2012; Goodman *et al.* 2012). This latter point is very important, because one recent study concluded that such biases are particularly likely in brief questionnaires such as the SDQ, which allow no role for clinical judgment (Goodman *et al.* 2012). The authors pointed out that due to these undesirable attributes, cross-national differences in SDQ caseness do not necessarily reflect comparable differences in disorder rates.

Beyond the inherent cross-cultural validity concerns, the extent to which differences in cross-national prevalence rates estimated by a self-report are determined by its measuring construct (i.e., factorial structure) is not clear so far. Actually, for a meaningful comparison across groups, demonstrating the measurement equivalence in the constructs underlying

one questionnaire across the groups is required (Gregorich, 2006; Milfont & Fisher, 2010). There appears to be a prevailing notion that the replicability of a factorial structure of one questionnaire in different cultural groups guarantees that the questionnaire will operate equivalently across these groups and it is suitable for cross-cultural comparisons (e.g., Byrne & Watkins, 2003). However, a prerequisite for cross-cultural comparisons is that the same theoretical construct is measured in each culture in the same way, namely that construct equivalence is achieved for the questionnaire measuring the construct when tested simultaneously across several cultural groups (He & van de Vijver, 2012). This is known as measurement equivalence (i.e., invariance) (Horn & McArdle, 1992). Therefore, in order to compare estimates by the questionnaire across various nations/countries, an important aspect that needs to be demonstrated is that reproducible factorial structure across different ethnic/cultural groups is also invariant (e.g., Byrne & Watkins, 2003, Gregorich, 2006; Milfont & Fisher, 2010). Several types of measurement invariance form a nested hierarchy: dimensional, configural, metric, scalar and strict factorial (Byrne & Watkins, 2003; Gregorich, 2006). Dimensional invariance refers that the same number of common factors are present across groups. Assuming dimensional, configural invariance refers that the same items are associated with the same factors across groups. Assuming configural, metric invariance refers that the common factors have the same meaning across groups (i.e., the equivalence of factor loadings). Assuming metric, scalar invariance refers to the equivalent intercepts or threshold of the items and is required to compare latent means across groups. Strict factorial dictates that the regression residual variances for all items are equal across groups.

The most striking observation from validation studies of the SDQ self-report is that the replicability of the same factorial structure of the self-report is not achieved across different ethnic/cultural groups. Some factor analytic studies using different language versions did support the original five-factor model including emotional symptoms, conduct problems, hyperactivity-inattention, peer problems, prosocial behaviour (e.g., Ronning *et al.* 2004; Ruchkin *et al.* 2007; Van Roy *et al.* 2008; Giannakopoulos *et al.* 2009). Other studies supported a modified five-factor model with reverse-worded items cross-loading on the other factors, being removed or with added error correlations to the factors (e.g. Van Roy *et al.* 2008; van de Looij-Jansen *et al.* 2011; Essau *et al.* 2012). A model with four factors was also supported including emotional symptoms and peer problems, conduct problems, hyperactivity-inattention, prosocial behaviour

(e.g. van de Looij-Jansen *et al.* 2011) and a model with three factors including internalising and externalising problems, prosocial behaviour (e.g. Koskelainen *et al.* 2001; Riso *et al.* 2010). Finally, some factor analytic studies failed to provide or provided modest support to proposed SDQ self-report models (e.g., Mellor & Stokes, 2007; Percy *et al.* 2008).

Contrary to heterogeneous data about SDQ self-report factor structure, the factor structure found for the YSR is fairly consistent across different societies. Using data from 23 and then 44 different societies in confirmatory factor analyses (CFA), Ivanova *et al.* (2007) and Rescorla *et al.* (2012) respectively demonstrated a consistent eight-syndrome measurement model including for the YSR. The eight-syndrome domains include: anxious/depressed, withdrawn/depressed, somatic complaints, social problems, thought problems, attention problems, rule-breaking behaviour and aggressive behaviour.

Turning to the measurement invariance of the YSR and SDQ self-report across different ethnic/cultural groups, there are several important findings. van de Looij-Jansen *et al.* (2011) demonstrated all forms of measurement invariance for proposed factor-models with the Dutch SDQ self-report in native Dutch and ethnic groups of Surinamese, Antillean/Aruban, Moroccan, Turkish and Capeverdean adolescents. Using the original English version of the SDQ and German, Cypriot Greek, Swedish and Italian translations, Essau *et al.* (2012) tested measurement invariance among adolescents from five European countries. The fit indices indicated that both the five-factor and the three-factor models provided good fit for the whole sample achieving only configural invariance. Using the self-report Norwegian translation of the SDQ, across native Norwegian and ethnic groups of Pakistani, Iranian, Turkish, Somali and Vietnamese adolescents, Richter *et al.* (2011) however failed to demonstrate the measurement invariance of the original five-factor model. On the other hand, Verhulp *et al.* (2014) demonstrated the full measurement invariance of three internalising syndrome scales of the YSR across four ethnic groups including native Dutch, Surinamese, Turkish and Moroccan adolescents. On a different note, Lambert *et al.* (2007) used the German and Jamaican versions of the YSR to test the original model considering item-response theory. They demonstrated that some items exhibit different item functioning in the YSR and only partially supported its measurement invariance (Lambert *et al.* 2007).

As shortly reviewed, there are scarce and heterogeneous results about the reproducibility of factorial structure and the measurement invariance of the two self-reports in multicultural contexts. Three of four studies on measurement invariance were organised

in the same country considering only ethnic groups. To which extent the findings from these studies are generic to ethnic minority adolescents in their host nations/countries remains unclear. Another important limitation of these studies is in the use of the main language version without considering the cultural adaptations to that version for ethnic minorities. However, the cultural adaptation of a questionnaire is important for ensuring conceptual equivalence in the measurements with that questionnaire (Poortinga, 1989), in order to avoid possible over- or under-evaluations of mental health from different ethnic groups. The only study evaluating the measurement invariance of the self-report across several nations included developed European countries (Essau *et al.* 2012), what significantly limit the generalisability of the findings to undeveloped and developing nations across the world with different socioeconomic development or cultural approach to mental health. This might be indicative that invariant cross-cultural general measures hardly exist and cross-cultural comparisons might be justified only for items within a general psychopathological measure that are identified as invariant across cultures.

Therefore, an important question that needs to be examined is the reproducibility of factorial structure of a self-report across different ethnic/cultural groups in order to evaluate whether different prevalence rates estimated by one questionnaire across various nations reflects true differences or the estimates are contaminated with the cultural-specific attributes related to the construct of interest. In order to provide more data on the applicability of the SDQ self-report measurement model in multicultural context, this study was organised to evaluate the measurement invariance of the SDQ self-report across seven national samples of convenience, sampled from India, Indonesia, Nigeria, Serbia, Turkey, Bulgaria and Croatia, which participate in our International Child Mental Health Study Group (ICMH-Study Group) project (Atilola, *et al.* 2013).

Methods

Participants

Data for the present study were obtained from the project organised by the ICMH-Study Group aiming to research mental health among children and adolescents living in undeveloped and developing countries (Atilola, *et al.* 2013). For the present study, data for adolescents aged 13–18 years from India, Indonesia, Nigeria, Serbia, Turkey, Bulgaria and Croatia were available. The same procedure was followed for recruiting participants in all countries. First, permission to

interview students was obtained from local authorities and/or appropriate ethical committees in each region. Afterwards, participants were sampled from the following regions of convenience: Kikinda, Belgrade and Zajecar in Serbia, Haerul Ihwan and Rizki Mulya Rahman in Indonesia, Ms Shelza in India, Ibadan in Nigeria, Primorsko-goranska in Croatia, Varna in Bulgaria and Sanliurfa in Turkey. From these regions two to five high schools in each country were randomly selected depending of number of pupils they had. The schools were randomly selected with a list of schools in the locality stratified where possible into rural or urban.

The sampling frame per country was 560 adolescents in the 9th to 12th grade. The participants were randomly contacted by the school psychologists or counsellors. They were selected by random picking (in no particular order) from the school register, taking cognisance of gender balance. The adolescents and their teachers were informed of the study by the school psychologists and investigators. Of all contacted, only those who agreed to participate and returned the written consents were included. The adolescents completed the ICMH-Study Group set of questionnaires at schools in order to prevent a low responding rate. The questionnaires were administered with the adolescents while seated in schools and they had enough space for comfort and privacy. To ensure those teachers and the school authorities had no insight into the adolescents' responses, teachers were excused from the hall and the adolescents were provided with sealable envelopes with which completed questionnaires were returned.

Instrument

The SDQ is a brief behavioural screening questionnaire for 3–16 year olds and it exists in several language versions (Goodman, 2001). Specifically, the SDQ self-report is available in more than 77 language versions translated and culturally validated following linguistic procedures provided by the developer (Goodman, 2001). The language versions used in the study were obtained from www.sdqinfo.org. The SDQ self-report has 25 items in five-item scales: emotional, conduct, hyperactivity/inattention and peer relationship problems and prosocial behaviour (Goodman, 2001). Each item has a three-point response scale (0 = not true; 1 = somewhat true; 2 = certainly true), with the five items of the problem scales that reflect strengths were reverse scored. The sum of all answered items in a scale creates its total score (possible range 0–10), while the sum of all answered items in the first four scales creates the total score (possible range 0–40).

Data analyses

A series of CFA was conducted to identify the best fitting model that can be applied in all countries. We tested seven different models available across the literature mentioned in the introduction. The original five-factor model, a three-factor model, a one-factor model, a bifactor model with five independent specific factors, a bifactor model with five correlating factor, a bifactor model with three independent specific factors, and finally a bifactor model with three correlating specific factors.

All analyses were performed with MPLUS 7.11 (Muthén & Muthén, 1998–2012). Weighted Least Squares Mean and Variance adjusted (WLSMV) estimation method was used (Brown, 2006; Finney & Di Stefano, 2006). The items were treated as ordinal indicators. The analyses were based on WLSMV estimation which utilises the entire weight matrix to compute s.e. for the parameters, but this method avoids the matrix inversion (Finney & Di Stefano, 2006).

In order to define the best fitting model we applied several fit indices. A satisfactory degree of fit requires the comparative fit index (CFI) and the Tucker–Lewis index (TLI) to be close to 0.95, and the model should be rejected when these indices are <0.90 (Brown, 2006). The next fit index was root-mean-squared error of approximation (RMSEA) and values below 0.05 indicate excellent fit, a value around 0.08 indicates adequate fit, and a value above 0.10 indicates poor fit (Browne & Cudek, 1993; Kline, 2011). Closeness of model fit using RMSEA (CFit of RMSEA) is a statistical test (Browne & Cudek, 1993), which evaluates the statistical deviation of RMSEA from the value 0.05. Non-significant probability values ($p > 0.05$) indicate acceptable model fit, though some methodologists would require larger values such as $p > 0.50$ (Brown, 2006). In order to compare alternative nested models using WLSMV estimator, we used the DIFFTEST procedure within MPLUS (Asparouhov & Muthén, 2006). We also planned to test the configural, metric and scalar invariance of SDQ, however we could not identify a measurement model that could be applied to data from all countries, therefore we did not perform these procedures.

After the failure of finding common measurement models across countries, we changed our strategies and performed exploratory factor analysis (EFA) in which we treated the indicators as ordinal scale, therefore the estimation method was also WLSMV and rotation was GEOMIN. In order to find the number of factors to extract, we also considered fit indices and interpretability of factor solutions. We considered important cross-loadings when the sizes of the factor loadings were higher than 0.30.

Table 1. Distribution of participants by age and gender across seven countries

Country, <i>n</i> (% response rate)	Gender*, male/female <i>n</i> (%)	<i>M</i> (s.d.) years**, age range
India, 393 (70.8)	244 (62.1)/149 (37.9)	14.60 (0.68), 13–16
Serbia, 386 (68.9)	173 (44.8)/213 (52.2)	16.68 (1.02), 15–18
Nigeria, 522 (93.2)	236 (45.2)/286 (54.8)	14.98 (1.26), 13–18
Turkey, 280 (50)	176 (62.9)/107 (37.1)	16.16 (0.95), 14–18
Indonesia, 228 (40.7)	105 (46.1)/123 (53.9)	16.13 (0.76), 14–18
Bulgaria, 265 (47.3)	129 (48.7)/136 (51.3)	15.33 (1.11), 14–17
Croatia, <i>n</i> = 293 (52.3)	121 (41.3)/172 (58.7)	16.19 (1.23), 14–18

* χ^2 (df) = 60.89 (6), $p < 0.001$; ** F (df) = 201.73 (6), $p < 0.0001$.

Results

Data from 2367 adolescents was available for this study. There were statistically significant differences between the countries in the participants' age ($p < 0.0001$) and gender ($p < 0.001$) (Table 1). Table 2 shows the SDQ scores across the countries.

Confirmatory factor analyses

Seven competing measurement models and the fit indices tested across the countries are presented in Table 3. The one-factor model (Model 1) was a starting model which yielded an inadequate level of fit in all countries. The original five-factor model (Model 2) also generated inadequate fit degree in all countries. Only in Croatia, the degree of fit of this model was close to the acceptable level, however in all other countries neither the RMSEA nor CFI, TLI were close to the acceptable. Because specification searches based on

modification indices are more likely to be successful when the model contains only minor misspecifications (MacCallum, 1986; Brown, 2006), we did not examine further the cross-loadings and error covariances in this model. The model depicting three correlating factors (Model 3) also did not reach the adequate degree of fit. The classical bifactor model (Model 4) which specifies one general factor and five uncorrelated specific factors did not fit the data satisfactorily, even in data from two countries this model could not be identified. We tested a bifactor model with one general factor and three uncorrelated specific factors (Model 5), but this model did not reach again the satisfactory degree of fit. We estimated two modified bifactor models which contains the correlations between specific factors. Model 6 which is a bifactor model with five correlating specific factors and yielded adequate degree of fit in data from four countries including India, Nigeria, Turkey and Croatia. In all other countries, the degree of fit approached the adequate level. We also tested a bifactor model with three correlating specific factors (Model 7) which also yielded adequate degree of fit in India, Nigeria, Turkey and Croatia. Comparison of Model 6 and Model 7 resulted significant $\Delta\chi^2$ value ranged between 20.0–44.1 (at least $p < 0.006$) with df=7 in six countries, $\Delta\chi^2$ value was not significant ($\Delta\chi^2=7.9$; df=7 $p=0.35$) only in the Indonesian sample. Due to the lack of common acceptable model or in other words lack of dimensional invariance for the seven countries it was not possible to perform the other types of the invariance test.

Exploratory factor analyses

Because the confirmatory analyses revealed that none of the tested models fits well to the data across the countries, we performed a series of EFA on data from each country separately in order to find which factors with corresponding items are replicable across the countries. EFA was performed with WLSMV estimator and we extracted five factors based on previous research and the inspections of eigenvalues and fit

Table 2. Distribution of the SDQ scores across seven countries

SDQ scale	India, <i>n</i> = 393	Serbia, <i>n</i> = 386	Nigeria, <i>n</i> = 522	Turkey, <i>n</i> = 280	Indonesia, <i>n</i> = 228	Bulgaria, <i>n</i> = 265	Croatia, <i>n</i> = 293
Emotional symptoms	2.9 (2.2)	2.7 (2.3)	4.2 (2.4)	3.3 (2.3)	4.1 (2.4)	2.5 (2.2)	2.4 (2.1)
Conduct symptoms	3.2 (1.7)	2.6 (1.6)	3.0 (2.2)	3.1 (1.9)	2.9 (1.6)	2.2 (1.7)	2.0 (1.3)
Hyperactivity symptoms	3.3 (1.9)	2.9 (2.1)	2.7 (2.0)	4.2 (1.8)	3.8 (1.7)	3.3 (2.0)	3.1 (2.0)
Peer problems	2.3 (1.5)	2.0 (1.5)	3.2 (2.0)	3.1 (1.6)	2.5 (1.4)	2.3 (1.7)	1.8 (1.6)
Prosocial	7.9 (1.9)	8.0 (1.8)	8.1 (2.0)	7.6 (2.0)	7.8 (1.6)	7.4 (2.0)	8.1 (1.7)
Total score	11.6 (5.2)	10.2 (5.1)	13.0 (6.0)	13.7 (5.2)	13.3 (4.6)	10.3 (4.9)	9.3 (4.8)

Table 3. Degree of model fit for five competing measurement models of the SDQ from seven different countries

		χ^2	df	RMSEA	Cfit of RMSEA	CFI	TLI
Model 1	<i>One-factor first-order model</i>						
	India	965.9	275	0.080	<0.001	0.533	0.490
	Serbia	1097.3	275	0.088	<0.001	0.520	0.477
	Nigeria	1090.9	275	0.075	<0.001	0.597	0.549
	Turkey	644.8	275	0.069	<0.001	0.703	0.676
	Indonesia	695.0	275	0.082	<0.001	0.439	0.388
	Bulgaria	1034.8	275	0.102	<0.001	0.385	0.329
	Croatia	721.0	275	0.074	<0.001	0.655	0.624
Model 2	<i>Five-factor first-order model</i>						
	India	606.8	265	0.057	0.023	0.769	0.738
	Serbia	651.4	265	0.061	<0.001	0.775	0.745
	Nigeria	911.5	265	0.068	<0.001	0.673	0.629
	Turkey	540.9	265	0.061	<0.001	0.778	0.749
	Indonesia	551.4	265	0.069	<0.001	0.618	0.567
	Bulgaria	584.6	265	0.067	<0.001	0.741	0.707
	Croatia	470.5	265	0.051	0.368	0.841	0.820
Model 3	<i>Three-factor first-order model</i>						
	India	633.2	272	0.058	0.012	0.756	0.731
	Serbia	734.5	272	0.067	<0.001	0.729	0.701
	Nigeria	1049.3	272	0.074	<0.001	0.606	0.566
	Turkey	580.0	272	0.064	0.001	0.753	0.727
	Indonesia	574.2	272	0.070	<0.001	0.597	0.555
	Bulgaria	682.5	272	0.075	<0.001	0.668	0.633
	Croatia	509.6	272	0.055	0.148	0.816	0.797
Model 4	<i>Bifactor model with uncorrelated five specific factors</i>						
	India	662.9	250	0.065	<0.0001	0.721	0.665
	Serbia	704.2	250	0.069	<0.0001	0.735	0.682
	Nigeria	916.3	250	0.071	<0.0001	0.663	0.595
	Turkey	516.4	250	0.062	0.006	0.786	0.743
	Indonesia	The model is not identified					
	Bulgaria	659.8	250	0.079	<0.0001	0.668	0.602
	Croatia	The model is not identified					
Model 5	<i>Bifactor model with uncorrelating three specific factors</i>						
	India	560.6	250	0.056	0.050	0.790	0.748
	Serbia	647.0	250	0.064	<0.001	0.768	0.722
	Nigeria	731.2	250	0.061	<0.001	0.756	0.708
	Turkey	The model is not identified					
	Indonesia	The model is not identified					
	Bulgaria	637.8	250	0.077	<0.001	0.686	0.623
	Croatia	454.6	250	0.053	0.266	0.842	0.810
Model 6	<i>Bifactor model with correlated five specific factors</i>						
	India	372.6	240	0.030	10.000	0.941	0.926
	Serbia	447.5	240	0.047	0.735	0.879	0.849
	Nigeria	379.6	240	0.033	10.000	0.929	0.912
	Turkey	331.6	240	0.037	0.991	0.926	0.908
	Indonesia	346.0	240	0.044	0.831	0.858	0.823
	Bulgaria	409.0	240	0.052	0.373	0.0.863	0.829
	Croatia	362.4	240	0.042	0.944	0.905	0.882
Model 7	<i>Bifactor model with correlated three specific factors</i>						
	India	367.3	247	0.035	10.000	0.919	0.901
	Serbia	480.9	247	0.050	0.537	0.864	0.834
	Nigeria	412.2	247	0.036	10.000	0.916	0.898
	Turkey	362.0	247	0.041	0.958	0.908	0.888
	Indonesia	355.8	247	0.044	0.837	0.855	0.824

Continued

Table 3. Continued

	χ^2	df	RMSEA	Cfit of RMSEA	CFI	TLI
Bulgaria	448.6	247	0.056	0.132	0.837	0.802
Croatia	383.2	247	0.043	0.903	0.895	0.872

Note. CFI, comparative fit index; TLI, Tucker–Lewis index; RMSEA, root-mean-squared error of approximation; Cfit of RMSEA, probability of RMSEA.

indices. Eigenvalues of factors in each sample are presented in Fig. 1. We also applied GEOMIN rotation which is an oblique type of rotation (Yates, 1987). Analysing the factor loadings we identified three common factors (see Appendix 1 online). The three factors were common in each country namely prosocial behaviour, emotional symptoms and conduct problems. We also found two factors that had different meanings in each country. After inspection of factor loadings, three items defined prosocial behaviour factor across countries (item 1 ‘try to be nice to other people’, 4 ‘share with others’ and 9 ‘being helpful’). Some other items also loaded considerably on this factor but not in all countries, in these latter cases these items loaded saliently on other factors. Emotional symptoms factor was formed by five items (item 3 ‘headaches, stomach-aches or sickness’, 8 ‘worry a lot’, 13 ‘unhappy’, 16 ‘being nervous in new situations’ and 24 ‘having many fears’), however all items have one

or two salient cross-loadings (>0.30) on other factors as well in some countries. Conduct problems factor was defined by two items including item 12 ‘fight a lot’, and 18 ‘accused of lying’.

Discussion

Previous studies with different language versions provided evidence to support different models for the SDQ self-report. However, a few studies tested the measurement models across various cultures and several countries. In the present report, we initially tested seven competing measurement models for the SDQ self-report in each country separately. Our results indicate that the original five-factor model had inadequate fit degree across all countries, contrary to the previous findings that mostly supported this model (e.g., Ronning *et al.* 2004; Ruchkin *et al.* 2007; Van Roy

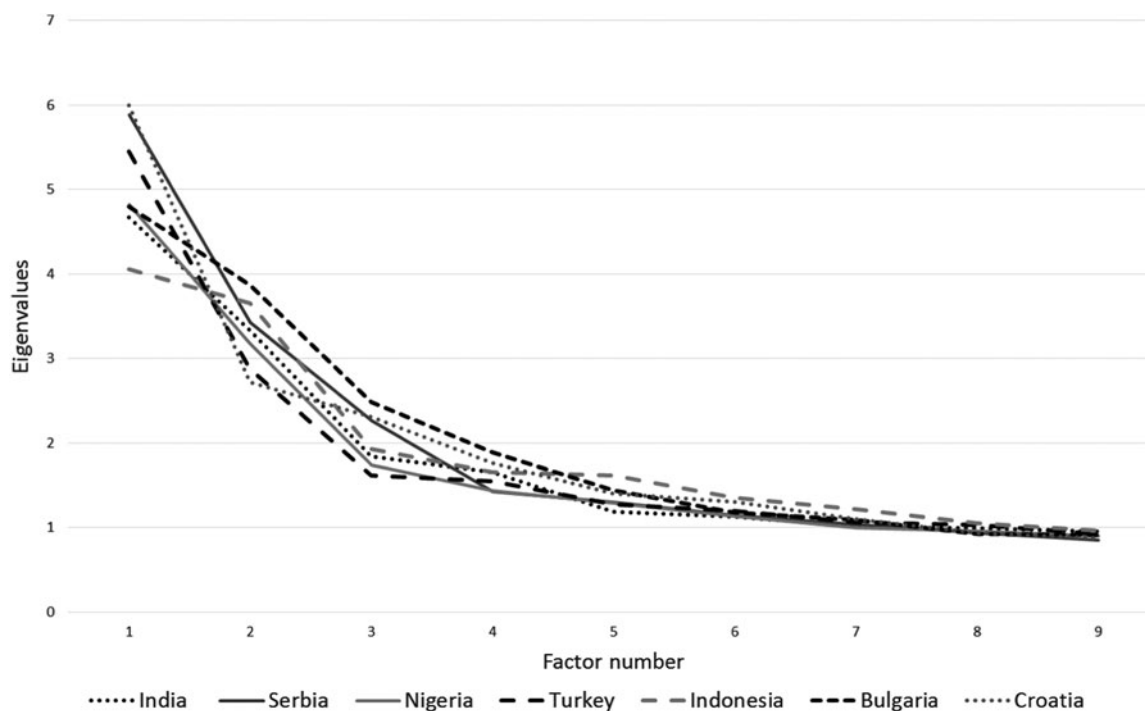


Fig. 1. Eigenvalues of factors of EFA.

et al. 2008; Giannakopoulos *et al.* 2009). Additionally, the model depicting three correlating factors did not reach the adequate degree of fit, contrary to some previous studies that reported acceptable model fit (e.g., Koskelainen *et al.* 2001; Dickey & Blumberg, 2004; Riso *et al.* 2010). Furthermore, the classical bifactor model which specifies one general factor and five uncorrelated specific factors did not fit the data satisfactorily as well, even in data from Indonesia and Croatia could not be identified. Considering that the most recent study supported a modification as a bifactor model including the five-factors and the general problem factor (Kóbor *et al.* 2013), we included this model and its modifications as well. A modified, bifactor model with six correlating specific factors and a bifactor model with three correlating specific factors yielded adequate fit degree in India, Nigeria, Turkey and Croatia, with the second one being more appropriate. This finding implies that the same five-factors and the general problem factor are common only for four countries. However, due to the lack of a common acceptable model across all seven countries, namely the same numbers of factors (i.e., dimensional invariance), it was not possible to perform the metric and scalar invariance test, which indicates that the SDQ self-report models tested lack appropriate measurement invariance across the countries included.

Turning to the results from a series of EFA on data from each country separately, it was observed that the prosocial behaviour, emotional symptoms and conduct problems factor were common for all countries. However, originally proposed items from these factors/scales loaded saliently on other factors besides the proposed ones or only some of them corresponded to proposed factors in all seven countries. Three items defining the prosocial behaviour factor were common for all countries, while the emotional symptoms factor was formed by five items. However, all items have one or two salient cross-loadings on the other factors. The conduct problems factor was defined only by two items. These items that loaded on the mentioned factors only could be regarded as culture-independent in the self-report. However, other items, especially the items of the peer problems and hyperactivity factors were perceived differently across the countries and they could be regarded as strongly influenced by specific factors – culture-dependent items.

A recent study using the data from Germany, Cyprus, England, Sweden, Italy tested the measurement invariance of the five-factor and three-factor model (Essau *et al.* 2012). A good fit of the data was found for the whole sample for both models, but it was observed that the SDQ structure might be different across the five countries. The study also confirmed only configural invariance, which has been found to be

an insufficient form of invariance for appropriate cross-cultural comparisons (Gregorich, 2006). The findings of our study strongly agree with this one about the measurement non-invariance of the SDQ self-report measurement model across nations indicating that the current SDQ models might not be suitable for comparisons in a multinational cross-cultural context. There may be several reasons for the measurement non-invariance of the SDQ self-report. First, there might be genuine differences in valuating, reporting and/or expressing psychological symptoms among adolescents from different nations, as it was observed for the YSR (Lambert *et al.* 2007). Accumulated evidence on child and adolescent psychopathology shows significant variations in rates of disorders across socio-cultural/ethnic groups, with culture-specific mental disorders suspected, different manifestations of disorders, and levels of similarities or differences in risk factors across groups (e.g., Nazroo, 1998; Achenbach, *et al.* 2008; Nikapota & Rutter, 2008). Considering the SDQ is designed to screen for universally represented symptoms of specific disorders, it is possible that its items are more sensitive to one culture and less to another or that they are easily confounded by the culture-specific attributes related to the construct. In other words, the norms associated with a particular dimension in one culture confound cross-cultural comparisons. In this regard, some items might not represent specific psychopathology as intended or there might be some unimportant items compared to culture specific and reference norms (e.g. Heine *et al.* 2002), what was only possible to recognise during the translation and cultural adaptation process of the SDQ from English into other languages (Berry *et al.* 2002). Additionally, Goodman *et al.* (2010) observed that some labels like conduct problems and hyperactivity may be misleading when applied in general population samples, and this could also be a source of the observed difference. For example, if we suspect that some SDQ factors such as conduct problem and attention deficit/hyperactivity problem share common factors (e.g. Heine *et al.* 2002), which might be perceived and operated in different ways cross-culturally. One important thing is that considering high percentages of comorbidity in symptoms in child and adolescents psychopathology, there might be some unrecognised, underlying impact on how the SDQs are clustered in different cultures. Last, considering that we tested only adolescents, there might be less non-invariance if parent or teacher report is used, which need to be explored in future studies using a multitrait-multimethod type of analysis which requires administering all three reports of the SDQ.

Our findings have several research implications. The findings imply that the current SDQ self-report

measurement model might not allow direct cross-country comparisons in levels of adolescent psychopathology. This finding, if further replicated, has implication for epidemiological and clinical research. In evaluating the impact of interventions targeted at improving the general mental health of children and adolescents in multinational context; researchers may be cautious in using the SDQ as a measure of pre/post-intervention changes in mental health. More clinically oriented measures may be more useful if not outright clinical evaluation itself. This note of caution is even more apt at this time when the impact of the sundry Millennium Development Goal (MDG) interventions may soon be evaluated, and child mental health may be considered as one of the outcome measures. This does not imply that it cannot be used for in-country comparisons when the specific norms are developed for that country. Currently available alternative to the SDQ self-report for cross-cultural comparative research could be the YRS, however, sufficient evidence for its strong measurement invariance also lacks. Consequently, cross-cultural comparisons might be justified for SDQ items identified as invariant across cultures or the SDQ self-report needs to be revised for meaningful cross-cultural comparisons. Possible revisions to the SDQ need primarily to be based on creating items that are more culture-independent and less culture-dependent, which can be easily implemented in multicultural contexts. This is probably achievable when future research attempt to contrast the weaker items found in the present study with open-ended questions or, preferably, interviews to validate their meaning, understanding and rating across different cultures. Furthermore, implementing minor model modifications, changing response format or items rewording suggested previously would be of importance as well (e.g. Ronning *et al.* 2004; Giannakopoulos *et al.* 2009; Essau *et al.* 2012; Kóbor *et al.* 2013). It has also been suggested that symptom ratings may achieve better cross-cultural comparability when assessments were based on more objective measures in which persons compare them with a certain standard instead of rating symptoms on scales like Likert scales (e.g. Heine *et al.* 2002). This standard could be an arithmetic mean for a particular SDQ scale.

There are some limitations that need to be taken into consideration when interpreting our findings. First, there are significant differences in the participant's age and gender distribution across the countries, what could bias the findings considering that some items might be more or less sensitive or important to age and gender than others in a specific nation. Additionally, only adolescents who agreed to participate were included and the response rate varied

substantially between the countries. Second, participants were sampled from regions of convenience, although schools in the regions were randomly selected, what could limit generalisability of the findings to adolescents from other country's regions. Additionally, although making sure to include samples with different socioeconomical, culturally and religious backgrounds, this does not imply that the seven countries included are representative of the developing and undeveloped world, and the generalisability of our finding is further limited. Third, the data were based solely on the adolescents' self-report, which dichotomise the outcome, and no behavioural observations or clinical indices were used to confirm this self-report measure (Purgato & Barbui, 2012). Fourth, parent and teacher report were also not tested for invariance and clinical samples were not included. To improve the robustness of findings, future studies evaluating SDQ models in the multicultural context and attempting to revise the measurement model need to be based on all three SDQ reports and may do well to include clinical samples.

In conclusion, the study showed the measurement non-invariance of the SDQ self-report measurement model across several nations indicating that the current SDQ models might not be suitable for cross-national cross-cultural comparisons.

Acknowledgements

We would like to thank to all adolescents who participated in this project. In addition, we would like to thank to two reviewers who gave substantial comments to the manuscript.

Financial Support

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Conflict of Interests

None.

Ethical Standards

All adolescents were informed about the aims and procedures of the study. Those who were 16 years and above signed consent forms while the younger participants returned signed parental consent and personal assent forms. Additionally, the authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in

2008. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional guides on the care and use of laboratory animals.

Supplementary material

The supplementary materials referred to in this article can be found at <http://dx.doi.org/10.1017/S2045796014000201>

References

- Achenbach TM** (1991a). *Manual for the Child Behavior Checklist/4–18 and 1991 Profile*. University of Vermont, Department of Psychiatry: Burlington, VT.
- Achenbach TM** (1991b). *Manual for the Teacher's Report Form and 1991 Profile*. University of Vermont, Department of Psychiatry: Burlington, VT.
- Achenbach TM** (1991c). *Manual for the Youth Self-Report and 1991 Profile*. University of Vermont, Department of Psychiatry: Burlington, VT.
- Achenbach TM, Rescorla LA** (2007). *Multicultural Supplement to the Manual for the ASEBA School-Age Forms & Profiles*. University of Vermont Research Center for Children, Youth, and Families: Burlington, VT.
- Achenbach TM, Becker A, Döpfner M, Heiervang E, Roessner V, Steinhausen HC, Rothenberger A** (2008). Multicultural assessment of child and adolescent psychopathology with ASEBA and SDQ instruments: research findings, applications, and future directions. *Journal of Child Psychology and Psychiatry* **49**, 251–275.
- Achenbach TM, Rescorla LA, Ivanova MY** (2012). International epidemiology of child and adolescent psychopathology I: diagnoses, dimensions, and conceptual issues. *Journal of the American Academy of Child and Adolescent Psychiatry* **51**, 1261–1272.
- Asparouhov T, Muthén B** (2006). Robust Chi square difference testing with mean and variance adjusted test statistics. Mplus Web Notes, 10. Retrieved 20 March 2014 (<http://www.statmodel.com/download/webnotes/webnote10.pdf>).
- Atilola O, Balhara YPS, Stevanovic D, Avicenna M, Kandemir H** (2013). Self-reported mental health problems among adolescents in developing countries: results from an international pilot sample. *Journal of Developmental and Behavioral Pediatrics* **34**, 129–137.
- Berry JW, Poortinga YH, Segall MH, Dasen PR** (2002). Methodological concerns. In *Cross-cultural Psychology: Research and Applications* (ed. JW Berry, YH Poortinga, MH Segall and PR Dasen), pp. 286–316. Cambridge University Press: Cambridge.
- Brown TA** (2006). *Confirmatory Factor Analysis for Applied Research*. Guilford Press: New York.
- Browne MV, Cudek R** (1993). Alternative ways of assessing model fit. In *Testing Structural Equation Models* (ed. KA Bollen and SJ Long), pp. 136–162. Sage: Newbury Park, CA.
- Byrne BM, Watkins D** (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology* **34**, 155–175.
- Camras LA, Fatani SS** (2006). The development of emotional expressivity and the influence of culture. *International Society for the Study of Behavioural Development Newsletter* **49**, 12–15.
- Dickey WC, Blumberg SJ** (2004). Revisiting the factor structure of the strengths and difficulties questionnaire: United States, 2001. *Journal of the American Academic Child Adolescent Psychiatry* **43**, 1159–1167.
- Essau CA, Olaya B, Anastassiou-Hadjicharalambous X, Pauli G, Gilvarry C, Bray D, O'callaghan J, Ollendick TH** (2012). Psychometric properties of the strength and difficulties questionnaire from five European countries. *International Journal of Methods in Psychiatric Research* **21**, 232–245.
- Finney SJ, Di Stefano C** (2006). Nonnormal and categorical data in structural equation modeling. In *Structural Equation Modeling: A Second Course* (ed. GR Hancock and RD Mueller), pp. 269–314. Information Age: Greenwich, CT.
- Giannakopoulos G, Tzavara C, Dimitrakaki C, Kolaitis G, Rotsika V, Tountas Y** (2009). The factor structure of the strengths and difficulties questionnaire (SDQ) in Greek adolescents. *Annals of General Psychiatry* **8**, 20.
- 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 Academy of Child and Adolescent Psychiatry* **40**, 1337–1345.
- Goodman R, Ford T, Corbin T, Meltzer H** (2004). Using the strengths and difficulties questionnaire (SDQ) multi-informant algorithm to screen looked after children for psychiatric disorders. *European Child and Adolescent Psychiatry* **13**, 25–31.
- Goodman A, Lamping DL, Ploubidis GB** (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.
- Goodman A, Heiervang E, Fleitlich-Bilyk B, Alyahri A, Patel V, Mullick MS, Slobodskaya H, Dos Santos DN, Goodman R** (2012). Cross-national differences in questionnaires do not necessarily reflect comparable differences in disorder prevalence. *Social Psychiatry and Psychiatric Epidemiology* **47**, 1321–1331.
- Greally P, Kelleher I, Murphy J, Cannon M** (2010). Assessment of the mental health of Irish adolescents in the community. *Royal College of Surgeons in Ireland Student Medical Journal* **3**, 33–35.
- Gregorich SE** (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care* **44**, S78–94.
- Hackett R, Hackett L** (1999). Child psychiatry across cultures. *International Review of Psychiatry* **11**, 225–235.

- He J, van de Vijver F (2012). Bias and equivalence in cross-cultural research. *Online Readings in Psychology and Culture* 2. <http://dx.doi.org/10.9707/2307-0919.1111>.
- Heiervang E, Goodman A, Goodman R (2008). The Nordic advantage in child mental health: separating health differences from reporting style in a cross-cultural comparison of psychopathology. *Journal of Child Psychology and Psychiatry* 49, 678–685.
- Heine SJ, Lehman DR, Peng K, Greenholtz J (2002). What's wrong with cross-cultural comparisons of subjective Likert scales?: The reference-group effect. *Journal of Personality and Social Psychology* 82, 903.
- Horn JL, McArdle JJ (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research* 18, 117–144.
- Ivanova MY, Achenbach TM, Rescorla LA, Dumenci L, Almqvist F, Bilenberg N, Bird H, Broberg AG, Dobrean A, Döpfner M, Erol N, Fornas M, Hannedottir H, Kanbayashi Y, Lambert MC, Leung P, Minaei A, Mulatu MS, Novik T, Oh KJ, Roussos A, Sawyer M, Simsek Z, Steinhausen HC, Weintraub S, Winkler Metzke C, Wolanczyk T, Zilber N, Zukauskienė R, Verhulst FC (2007). The generalizability of the youth self-report syndrome structure in 23 societies. *Journal of Consulting and Clinical Psychology* 75, 729–738.
- Kline RB (2011). *Principles and Practice of Structural Equation Modeling*, 3rd edn. Guilford Press: New York.
- Kóbor A, Takács Á, Urbán R (2013). The bifactor model of the strengths and difficulties questionnaire. *European Journal of Psychological Assessment* 29, 299–307.
- Koskelainen M, Sourander A, Vauras M (2001). Self-reported strengths and difficulties in a community sample of Finnish adolescents. *European Child and Adolescent Psychiatry* 10, 180–185.
- Lai KY, Luk ES, Leung PW, Wong AS, Law L, Ho K (2010). Validation of the Chinese version of the strengths and difficulties questionnaire in Hong Kong. *Social Psychiatry and Psychiatric Epidemiology* 45, 1179–1186.
- Lambert MC, Essau CA, Schmitt N, Samms-Vaughan ME (2007). Dimensionality and psychometric invariance of the youth self-report form of the child behavior checklist in cross-national settings. *Assessment* 14, 231–245.
- Mabe PA, Josephson AM (2004). Child and adolescent psychopathology: spiritual and religious perspectives. *Child and Adolescent Psychiatric Clinics of North America* 13, 111–125.
- MacCallum RC (1986). Specification searches in covariance structure modelling. *Psychological Bulletin* 100, 107–120.
- Mellor D, Stokes M (2007). The factor structure of the strengths and difficulties questionnaire. *European Journal of Psychological Assessment* 23, 105–112.
- Milfont TL, Fisher R (2010). Testing measurement invariance across groups: applications for cross-cultural research. *International Journal of Psychological Research* 3, 111–121.
- Muthén LK, Muthén BO (1998–2012). *Mplus User's Guide*, 7th edn. Muthén & Muthén: Los Angeles.
- Nazroo JY (1998). Genetic, cultural or socio-economic vulnerability? Explaining ethnic inequalities in health. *Sociology of Health and Illness* 20, 710–730.
- Nikapota A, Rutter M (2008). Sociocultural/ethnic groups and psychopathology. In *Rutter's Child and Adolescent Psychiatry*, 5th edn (ed. MD Rutter, VM Bishop, DS Pine, S Scott, J Stevenson, E Taylor, A Thapar), pp. 199–211. Blackwell Publishing Limited: London.
- 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.
- Poortinga YH (1989). Equivalence of cross-cultural data: an overview of basic issues. *International Journal of Psychology* 24, 737–756.
- Purgato M, Barbu C (2012). Dichotomizing rating scale scores in psychiatry: a bad idea. *Epidemiology and Psychiatric Science*, 22, 17–19.
- Ravens-Sieberer U, Wille N, Erhart M, Bettge S, Wittchen HU, Rothenberger A, Herpertz-Dahlmann B, Resch F, Hölling H, Bullinger M, Barkmann C, Schulte-Markwort M, Döpfner M (2008a). Prevalence of mental health problems among children and adolescents in Germany: results of the BELLA study within the national health interview and examination survey. *European Child and Adolescent Psychiatry* 17, 22–33.
- Ravens-Sieberer U, Erhart M, Gosch A, Wille N (2008b). Mental health of children and adolescents in 12 European countries – results from the European KIDSCREEN study. *Clinical Psychology and Psychotherapy* 15, 154–163.
- Rescorla L, Ivanova MY, Achenbach TM, Begovac I, Chahed M, Drugli MB, Emerich DR, Fung DS, Haider M, Hansson K, Hewitt N, Jaimes S, Larsson B, Maggolini A, Marković J, Mitrović D, Moreira P, Oliveira JT, Olsson M, Ooi YP, Petot D, Pisa C, Pomalima R, da Rocha MM, Rudan V, Sekulić S, Shahini M, de Mattos Silveiras EF, Szivovics L, Valverde J, Vera LA, Villa MC, Viola L, Woo BS, Zhang EY (2012). International epidemiology of child and adolescent psychopathology. II. Integration and applications of dimensional findings from 44 societies. *Journal of the American Academy of Child & Adolescent Psychiatry* 51, 1273–1283.
- Richter J, Sagatun Å, Heyerdahl S, Oppedal B, Røysamb E (2011). The strengths and difficulties questionnaire (SDQ) – self-report. An analysis of its structure in a multiethnic urban adolescent sample. *Journal of Child Psychology and Psychiatry* 52, 1002–1011.
- Riso DD, Salcuni S, Chessa D, Raudino A, Lis A, Altoè G (2010). The strengths and difficulties questionnaire (SDQ). Early evidence of its reliability and validity in a community sample of Italian children. *Personality and Individual Differences* 49, 570–575.
- Ronning JA, Handegaard BH, Sourander A, Mørch WT (2004). The strengths and difficulties self-report questionnaire as a screening instrument in Norwegian community samples. *European Child and Adolescent Psychiatry* 13, 73–82.
- Ruchkin V, Kuposov R, Schwab-Stone M (2007). The strength and difficulties questionnaire: scale validation with Russian adolescents. *Journal of Clinical Psychology* 63, 861–869.
- Thabet AA, Stretch D, Vostanis P (2000). Child mental health problems in Arab children: application of the strengths and

- difficulties questionnaire. *International Journal of Social Psychiatry* **46**, 266–280.
- van de Looij-Jansen PM, Goedhart AW, de Wilde EJ, Treffers PD** (2011). Confirmatory factor analysis and factorial invariance analysis of the adolescent self-report strengths and difficulties questionnaire: how important are method effects and minor factors? *British Journal of Clinical Psychology* **50**, 127–144.
- Van Roy B, Veenstra M, Clench-Aas J** (2008). Construct validity of the five-factor strengths and difficulties questionnaire (SDQ) in pre, early, and late adolescence. *Journal of Child Psychology and Psychiatry* **49**, 1304–1312.
- Verhulp EE, Stevens GWJM, Van de Schoot R, Vollebergh WAM** (2014). Using the youth self-report internalizing syndrome scales among immigrant adolescents: testing measurement invariance across groups and over time. *European Journal of Developmental Psychology* **11**, 102–110.
- Yates A** (1987). *Multivariate Exploratory Data Analysis: A Perspective on Exploratory Factor Analysis*. State University of New York Press: Albany.