The Bifactor Model of the Strengths and Difficulties Questionnaire

Andrea Kóbor[a,b,c] (kobor@cogpsyphy.hu)

Ádám Takács[a,b,c] (takacs@cogpsyphy.hu)

Róbert Urbán[b] (urban.robert@ppk.elte.hu)

[a]: Doctoral School of Psychology, Eötvös Loránd University; Izabella utca 46., H-1064, Budapest, Hungary

[b]: Institute of Psychology, Eötvös Loránd University; Izabella utca 46., H-1064, Budapest, Hungary

[c]: Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences; Szondi utca 83-85., H-1068, Budapest, Hungary

Manuscript of the article that appeared in:

European Journal of Psychological Assessment 2013; Vol. 29(4):299–307

DOI: 10.1027/1015-5759/a000160

Author Note

Correspondence concerning this article should be addressed to Andrea Kóbor, Institute of Cognitive Neuroscience and Psychology, Research Centre for Natural Sciences, Hungarian Academy of Sciences; Szondi utca 83-85., H-1068, Budapest, Hungary.

E-mail: kobor@cogpsyphy.hu

Summary

Background: The Strengths and Difficulties Questionnaire (SDQ) is a frequently used instrument developed for screening childhood mental health problems. *Aims*: The aim of this study is to clarify the structure of the Hungarian version of SDQ, to test previous measurement models, and to propose an alternative bifactor model. *Methods*: Data were collected from a community sample of 8-13-year-old children. We conducted a series of confirmatory factor analyses on parent (n = 383) and teacher ratings (n = 391). The classic five-factor, an alternative three-factor, and a bifactor model were estimated. In the bifactor model, specific components refer to the five SDQ-traits, and the General Problems factor refers to an impression about the problem severity of the child. *Results*: For both informants, the bifactor model yielded the best fit to the data compared to other models. *Discussion*: Childhood behavioral problems can be best described as a multidimensional construct, which has implications regarding the screening procedure in various samples.

Keywords: behavioral problems, bifactor structure, parent and teacher report, screening, Strengths and Difficulties Questionnaire

Introduction

Mental health problems in childhood can have detrimental impact on psychological health and productivity in adult life (Stansfeld, Clark, Caldwell, Rodgers, & Power, 2008). Screening mental health in childhood can offer the opportunity to provide early treatment and prevention of later problems.

The Strengths and Difficulties Questionnaire (SDQ) is a widely used tool for identifying child behavioral problems. This brief questionnaire is also suitable for rapid screening of population and to detect atypicalities in the individual profiles (Achenbach et al., 2008; Warnick, Bracken, & Kasl, 2008). Developing short and well functioning screening tools is of urgent need due to the increasing prevalence of child and adolescent mental health disorders over the last 50 years (Hagquist, 2007). The SDQ has a theoretical five-scale structure including emotional symptoms (anxiety and depression), conduct problems (conduct disorder [CD] and oppositional defiant disorder [ODD]), hyperactivity/inattention (attention deficit/hyperactivity disorder [ADHD]), peer relationship problems, and prosocial behavior (R. Goodman, 1997). The content and the structure of the SDQ were developed with reference to the main categories of child mental health disorders described by the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; American Psychiatric Association, 2000; A. Goodman, Lamping, & Ploubidis, 2010).

Clinicians and researchers should assess the degree to which a child psychopathology is manifested (i.e., symptom severity) aside from the presence or absence of the disorder. Inclusion of a quantitative axis to the categorical taxonomy of the DSM could help with taking different sources of variance (affecting the expression of symptoms) into account (Hudziak, Achenbach, Althoff, & Pine, 2007). The reports of multiple informants could be essential to confirming impairments in multiple settings (e.g., at home and at school), and resolving the complex issue of the informant effect appears to be a goal of the DSM-5 (for the specific case of ADHD, see Nigg, Tannock, & Rohde, 2010; Valo & Tannock, 2010). Co-morbidity among adult and developmental psychopathologies is rather the rule, and symptomatically homogeneous groups are rarely found in community and in clinical samples (Hudziak, et al., 2007; Krueger, McGue, & Iacono, 2001).

The original five-factor structure of SDQ has been validated by several exploratory factor analytic studies (Becker, Hagenberg, Roessner, Woerner, & Rothenberger, 2004; R. Goodman, 1997). However, model-based confirmatory factor analytic (CFA) studies of SDQ have yielded contradictory results across different populations (A. Goodman, et al., 2010; Van Roy, Veenstra, & Clench-Aas, 2008), and a few reports have supported this solution (Giannakopoulos et al., 2009; Ronning, Handegaard, Sourander, & Morch, 2004; Ruchkin, Koposov, & Schwab-Stone, 2007; Van Roy, et al., 2008). In one study, the authors tested a modified version adding a “positive construal method factor” to the original model (Van Roy, et al., 2008). This factor captured the method effect of positively worded (reversed) items. An alternative theoretical approach proposes a three-factor structure: Internalizing (Emotional Symptoms and Peer Problems), Externalizing (Conduct Problems and Hyperactivity/Inattention), and Prosocial Behavior. This model was supported in three studies (Dickey & Blumberg, 2004; Koskelainen, Sourander, & Vauras, 2001; Riso et al., 2010). Goodman et al. (2010) proposed a more complex five-factor model in which Internalizing and Externalizing Problems are two second order factors beyond the four problem scales. This structure has been validated on a large sample using all the three versions of the questionnaire (parent, teacher, and youth/self-report).

Among complex factor models it is worth taking into account a more sophisticated bifactor model (Holzinger & Swineford, 1937), which allows a kind of “g” factor in addition to distinct factors. Bifactor structures are rarely applied in personality and health assessment, but there is an increasing agreement that psychiatric symptoms and disorders maintain hierarchical structure where general (common) and domain-specific (conceptually narrow) or unique components play important roles (e.g., Thomas, 2012). This model could provide an alternative to non-hierarchical multidimensional representations of individual differences (Reise, Morizot, & Hays, 2007), and it is suggested to be an effective approach to modeling construct-relevant multidimensionality (Reise, in press). The model confronts the fact that psychological constructs are hierarchical and that content diversity does not necessarily have to be partitioned out into redundant subscales (Reise, in press; Reise, Moore, & Haviland, 2010). In the typical specification of the bifactor model there are at least three indicators for each specific/group factor, and all factors are uncorrelated, i.e., they are in the same order or conceptual footing (Reise, et al., 2010; Yung, Thissen, & McLeod, 1999). However, correlations among specific factors could be identified (Reise, in press).

Indicators of childhood behavioral problems are divers, and heterogeneous content needs to be included in its measure. These problems can be described as hierarchical or multidimensional constructs with a general component and with more specific components on a lower level (Gibbons & Hedeker, 1992). The bifactor model allows for the indicators of “difficulties” or “strengths” to simultaneously load on an overall primary dimension and to have a secondary loading on a distinct factor of a different behavioral sub-domain. The specific factors model the residual association between the items once the contribution of the primary factor has been accounted for (Gibbons et al., 2007). Severity and transparency of day-to-day problems and difficulties in behavior regulation reported by teachers or parents in the presence of neuropsychological syndromes (or at risk of them) could be captured by a general dimension. Besides confirming multidimensionality, a general factor may also help to explain the high levels of co-morbidity (Rhee, Willcutt, Hartman, Pennington, & DeFries, 2008) between several fronto-striatal syndromes (ADHD, obsessive-compulsive disorder [OCD], CD/ODD, Anxiety, Depression), which are incorporated in the four problem scales of

SDQ. In a recent study, compared to one-factor, two-factor, three-factor, and second order factor models of disruptive behaviors a bifactor model provided the best level of fit to the symptoms of ADHD and ODD (Martel, Gremillion, Roberts, von Eye, & Nigg, 2010).

As a consequence of multidimensionality, interpretation of obtained scale scores and their reliability estimates (such as coefficient alpha) becomes ambiguous. Individual differences on item scores reflect differences in two constructs (in the general and in specific factors). Reporting total score and subscale scores is not a satisfactory solution, because reliability coefficients for the subscales can be inflated by the variance due to the general factor (Reise, in press). In addition, subscales may have differential correlates with external variables; and they may appear to be reliable, but that reliability is a function of the general dimension, not the specific domain (Reise, et al., 2010). Moreover, subscales are often unreliable compared to composite score indicating that the latter is a better predictor of an individual's true score on a subscale (Sinharay, Haberman, & Puhan, 2007).

According to our knowledge, a bifactor model of SDQ has not been estimated so far, moreover comparing alternative models of SDQ is still lacking in this field. The main goal of this study is to compare alternative measurement models of both parent and teacher reports of the questionnaire in order to clarify the psychometric properties and inner structure of SDQ. In this phase we do not intend to establish normative scores or to test construct/predictive validity of this screening tool.

Materials and Methods

Participants and Procedure

Data were collected from primary school students from the third to the fifth grades (8 to 13 years). Parents of 383 children (185 boys and 198 girls, M = 122.74 months; SD = 10.84 months), and teachers of 391 children (198 boys, 193 girls, M = 124.44 months; SD = 11 months) completed the corresponding versions of the questionnaire; but we could not obtain reports on every child from both informants. The total sample was composed of four schools, from which altogether 22 classes participated (one teacher rated 5 to 28 children), all situated in different districts of Budapest. A smaller part of the parental reports (105 questionnaires) were collected in an earlier stage of the project which involved rural schools as well. The presence and severity of certain child psychiatric disorders is unknown in such a non-clinical sample.

Our study was approved by the Institutional Review Board of Eötvös Loránd University, Hungary. The schools were informed about the aim of the research in person as well as in writing. The teacher reports were given to the class teachers concerned, and the parent reports were taken home by the children. The final participation rate in the present study was 76% for teachers and 62% for parents.

Measures

The Hungarian versions of SDQ (with impact supplement) for parents and for teachers were used (translated into Hungarian by Judit Gervai and Mária Székely, for details and free download, see www.sdqinfo.org). Both versions of the questionnaire consist of 25 items, which can be answered on three-point response scales (“Not true”; “Somewhat true”; “Certainly true”). In the case of five (“Obedient”, “Has good friend”, “Generally liked”, “Thinks before acting”, and “Good attention”) of the ten positively worded items the inverse scores were used in the analyses, but in regards to the five items of the Prosocial Scale, the original values were reckoned in.

Data analyses

In order to test the factor structure of the Hungarian SDQ, a series of confirmatory factor analyses (CFA) were conducted on parent and teacher datasets separately. These statistical analyses were performed with the MPLUS 6.1 program. We treated the items as ordinal indicators and used weighted least squares mean and variance adjusted estimation method (WLSMV; Brown, 2006; Finney & DiStefano, 2006). To account for the hierarchical nature of the teacher data (i.e., students were nested-within classes) corrections to the standard errors and chi-square test statistics of model fit were made to take into account the non-independence of observations. The number of missing data was minimal in the current study, affecting approximately .3 and .1 percent of the sample based on parent and teacher report.

We report multiple fit indices according to the usual practice; starting with the comparative fit index (CFI) and the Tucker–Lewis Index (TLI). To consider a model as showing a satisfactory degree of fit, we require these to be close to .95, and the model should be rejected when these indices are below .90 (Brown, 2006). The other fit index is root mean squared error of approximation (RMSEA). RMSEA below .05 indicates an excellent fit, a value around .08 indicates an acceptable fit, and a value above .10 indicates a poor fit. 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 .05. Non-significant probability values ($p > .05$) indicate an acceptable fit. We used the DIFFTEST procedure within MPLUS (Asparouhov & Muthén, 2006) to compare alternative nested models using WLSMV estimator.

We had to drop out one item ("Steals from home, school or elsewhere") from the confirmatory factor analyses due to the extremely skewed distribution showing that the relative frequency of "Somewhat true" response was only 1.3% in the parent report and 0.5% in the teacher report, while all the other responses were "Not true".

Results

Alternative measurement models

Our study tested four competing factor models for parent and teacher ratings separately. The models and the fit indices are presented in Table 1; visual representations of the models are shown in Figure 1. The one-factor model (Model 0) was a starting model which yielded an inadequate level of fit in both informants.

The classic first order model of the five freely correlating SDQ factors (Model 1) indicated an acceptable fit for the two informants, although RSMEA deviates from .05 according to closeness of model fit in both models. All factor loadings and correlations were significant in case of both informants. In our second model (Model 2), we treated Behavioral Problems and Hyperactivity as part of the broader Externalizing Problems, and Emotional Symptoms and Peer Problems as part of Internalizing Problems on the basis of high significant correlations between these pairs of first order latent scores (.81 and .51–.57), and based on the models suggested by Goodman et al. (2010). However, we conducted these analyses by replacing the above-mentioned four factors with first order Internalizing and Externalizing latent variables. The original Prosocial factor was kept, and we allowed these three factors to correlate freely, because the correlation coefficients indicated moderate to strong pair-wise relations between them. This more restrictive model exhibited a poor fit compared to the five-factor first order model, and according to CFI and TLI it should be rejected. All correlations and factor loadings were significant in the case of both informants.

Finally, we tested the bifactor model (Model 3) of the questionnaire. Besides theoretical aspects, as an empirical support, we found moderate to high correlations between the five SDQ traits, except between Prosocial Scale and Emotional Symptoms Scale (-.25 and -.16 for parent and teacher report), and between Behavioral Problems and Emotional Symptoms Scale in the teacher model (.18). Regarding all the relevant fit indices the bifactor model exhibited an excellent fit on parent and teacher data. Although the χ^2 statistics remained significant, the lowest values were obtained for the bifactor structure. The fit of the bifactor models were significantly better than the fit of the other five-factor ($\Delta\chi^2(24) = 167.7$, $p < .001$ for parent report; $\Delta\chi^2(24) = 246$, $p < .001$ for teacher report) or three-factor models ($\Delta\chi^2(31) = 288.9$, $p < .001$ for parent report; $\Delta\chi^2(31) = 466.4$, $p < .001$ for teacher report). We termed this general factor as “General Problems”, and it might be considered as an overall first impression about the child’s problem severity whose behavior is just under description. Hereby it could also imply to the transparency of the possible syndrome(s). Correlations between the general factor and each of the five specific factors were fixed at zero in both models in line with the usual specification. The vast majority of factor loadings of items on specific factors (see Table 2) remained salient in both models (>.30), supporting the notion that specific factors still explain further variance besides the general factor. In terms of the parent report, all the 24 indicators load significantly (all *p*s < .001) on the General Problems factor. Detailed analysis of the teachers’ bifactor structure showed somewhat different results. Five of the loadings (“Somatic symptoms”, “Solitary”, “Worries”, “Better with adults than with children”, “Many fears”) were not significant on the General Problems factor. The leading items were unambiguously related to hyperactive-impulsive and inattentive behavior (with loadings of .98–.71), and two other items concerning conduct problems (“Obedient”, .75; “Fights or bullies”, .72) were almost equally important. The General Problems factor explains the extremely large variance of items belonging to the Hyperactivity scale in the teacher sample. Nevertheless, two hyperactive items (“Restless” and “Fidgety”) did not have significant loadings on their original group factor. Teachers might consider these symptoms more as a general problem and less as specific mental difficulties. The three remaining hyperactive items also loaded highly on the General Problems factor. However, unexpectedly, they also loaded negatively on the residualized Hyperactivity factor which we understand as a product of the negative or cross-over suppression effect (Paulhus, Robins, Trzesniewski, & Tracy, 2004).

We allowed for the specific factors to freely correlate in the bifactor model, but the General Problems factor explained the correlations between the five SDQ traits. In the parent model, correlations between the residualized factors of Hyperactivity and Emotional Symptoms, Hyperactivity and Behavioral Problems, and Peer Problems and Behavioral Problems were no longer significant. A similar pattern occurred in the teacher model, where the correlations between Behavioral Problems and Emotional Symptoms, Hyperactivity and Behavioral Problems, and Prosocial Scale and Emotional Symptoms turned into non-significant.

Table 1: Degree of model fit for three competing measurement models of the Hungarian version of the Strengths and Difficulties Questionnaire for parents and teachers.

		χ^2	df	CFI	TLI	RMSEA	Cfit of RMSEA
Model 0	One-factor first order model						
	Parents	1040.0*	252	.785	.765	.090	< .001
	Teachers	2322.1*	252	.796	.776	.145	< .001
Model 1	Five-factor first order model						
	Parents	564.8*	242	.912	.900	.059	.010
	Teachers	856.7*	242	.939	.931	.081	< .001
Model 2	Three-factor first order model						
	Parents	736.1*	249	.867	.853	.071	< .001
	Teachers	1268.6*	249	.899	.889	.102	< .001
Model 3	*Five-factor bifactor model*						
	Parents	*365.3**	*218*	*.960*	*.949*	*.042*	.962
	Teachers	*462.3**	*218*	*.976*	*.970*	*.054*	.190

Note. CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean squared error of approximation;

Cfit of RMSEA = probability of RMSEA.

*$p < .001$

Model 0. One-factor first order model

Model 1. Five-factor first order model

Model 2. Three-factor first order model

Model 3. Five-factor bifactor model

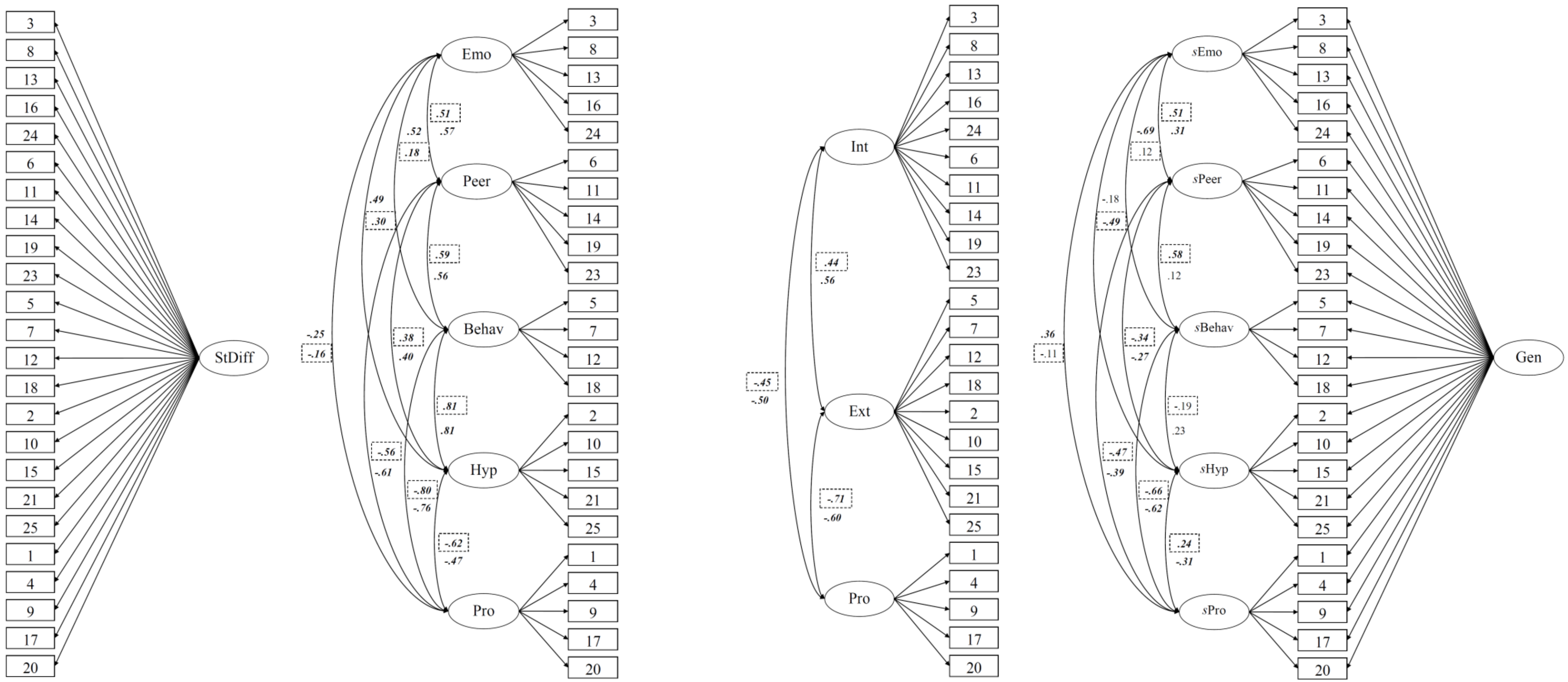

Figure 1

Figure 1

Schematic illustration of the four alternative models.

Note. Values is dashed boxes refer to correlations among factors in the teacher model, other values indicate correlations in the parent model. Significant correlations are boldfaced and italicized.

Model 0. One-factor first order model

Model 1. Five-factor first order model

Model 2. Three-factor first order model

Model 3. Five-factor bifactor model

Abbreviations for the factors: StDiff = Strengths and Difficulties (general factor in the one-factor model), Emo = Emotional Symptoms, Behav = Behavioral (Conduct) Problems, Hyp = Hyperactivity, Peer = Peer Problems, Pro = Prosocial Scale, Int = Internalizing Problems, Ext = Externalizing Problems, Gen = General Problems, *s*Emo = specific factor of Emotional Symptoms, *s*Behav = specific factor of Behavioral (Conduct) Problems, *s*Hyp = specific factor of Hyperactivity, *s*Peer = specific factor of Peer Problems, *s*Pro = specific factor of Prosocial Scale.

Abbreviations for the items: Somatic symptoms = 3, Worries = 8, Unhappy = 13, Nervous in new situations = 16, Many fears = 24, Solitary = 6, Has good friend = 11, Generally liked = 14, Picked on or bullied = 19, Better with adults = 23, Tempers = 5, Obedient = 7, Fights or bullies = 12, Lies or cheats = 18, Restless = 2, Fidgety = 10, Easily distracted = 15, Thinks before acting = 21, Good attention = 25, Considerate = 1, Shares readily = 4, Helpful if someone hurt = 9, Kind to younger children = 17, Often volunteers = 20.

Table 2: Fully standardized item loadings from bifactor models of parent and teacher SDQs.

Items	Emotional Symptoms		Behavioral Problems		Hyperactivity		Peer Problems		Prosocial Scale		General Problems	
	Parent	Teacher	Parent	Teacher	Parent	Teacher	Parent	Teacher	Parent	Teacher	Parent	Teacher
Somatic symptoms	**.34**	**.57**									**.31**	.08
Worries	**.68**	**.89**									**.34**	-.10
Unhappy	**.26**	**.72**									**.60**	**.28**
Nervous in new situations	**.59**	**.73**									**.44**	**.26**
Many fears	**.57**	**.84**									**.44**	-.06
Tempers			.02	**.38**							**.72**	**.66**
Obedient			**.45**	**.43**							**.61**	**.75**
Fights or bullies			**.44**	**.50**							**.59**	**.72**
Lies or cheats			.05	**.48**							**.74**	**.58**
Restless					**.74**	.14					**.56**	**.98**
Fidgety					**.67**	.12					**.61**	**.90**
Easily distracted					**.15**	**-.58**					**.75**	**.71**
Thinks before acting					**.15**	**-.31**					**.72**	**.84**
Good attention					.15	**-.62**					**.77**	**.62**
Solitary							**.60**	**.83**			**.26**	.03
Has good friend							**.63**	**.85**			**.32**	**.21**
Generally liked							**.63**	**.78**			**.59**	**.46**
Picked on or bullied							**.55**	**.57**			**.45**	**.42**
Better with adults							**.55**	**.75**			**.26**	.00
Considerate									**.52**	**.63**	**-.57**	**-.67**
Shares readily									**.50**	**.79**	**-.29**	**-.39**
Helpful if someone hurt									**.41**	**.75**	**-.32**	**-.53**
Kind to younger children									**.45**	**.62**	**-.32**	**-.53**
Often volunteers									**.40**	**.60**	**-.45**	**-.43**
Explained common variance (%)	11.0	17.5	3.3	4.9	8.9	5.2	14.6	17.6	8.8	14.1	53.3	40.7

Note. Significant factor loadings are boldfaced.

Table 3: Model-based scale score reliabilities for the bifactor models of parent and teacher SDQs.

	ω	ω_h	$\omega_{h(G)}$	ω	ω_h	$\omega_{h(G)}$
	Parent			Teacher		
Emotional Symptoms	.79	.45	.34	.88	.87	.01
Behavioral Problems	.81	.09	.72	.89	.27	.62
Hyperactivity	.91	.21	.70	.96	.08	.87
Peer Problems	.83	.59	.24	.90	.83	.07
Prosocial Scale	.74	.43	.31	.93	.59	.34

Note. ω = omega; ω_h = omega hierarchical; $\omega_{h(G)}$ = omega hierarchical related to General Problems factor. Values range from 0 (no reliability) to 1 (perfect reliability).

Factor reliabilities and indices of unidimensionality

In case of multidimensionality and hierarchically structured constructs, Cronbach's α is misleading in how well a measure reflects a single construct (Cortina, 1993). As we apply bifactor structures, model-based reliability estimates should be computed that denote how precisely a certain scale score assesses the combination of general and specific constructs, and a certain target construct (Brunner, Nagy, & Wilhelm, 2012, p. 831). To evaluate the measurement precision of each subscale in assessing the blend of General Problems and specific mental problems (e.g., Emotional Symptoms) we calculated coefficient omega; and in assessing only specific problems or only General Problems we computed coefficient omega hierarchical (for details, see Brunner, et al., 2012). Model-based score reliabilities are presented in Table 3. Both in parent and teacher datasets scale scores contained a large amount of variance attributable to the blend of general and specific problems (ω ranges from .74 to .96). The coefficients are slightly lower in the parent model (especially in case of the Prosocial Scale, which could be diverse as the related behaviors might be more transparent in school environment). Emotional Symptoms and Peer Problems subscales measured well the specific problem constructs in the teacher model (.87 and .83). This was not true for the parent model to such a degree (.45 and .59). Low values of omega hierarchical related to General Problems factor in the teacher model (.01 and .07) also indicated that these problems concerning internalization did not measure a general construct precisely. All coefficients in both models indicated that Hyperactivity and Behavioral Problems properly measure general mental difficulties.

A clearer index of unidimensionality could be to define the percent of common variance attributable to the general factor through the use of an explained common variance index (ECV, Bentler, 2009; Berge & Sočan, 2004). We estimated the ECV in both models, and found that the General Problems factor explains 53 % of common variance in parental ratings and 41 % in teachers' ratings. We also estimated the ECV values for each specific factor (see in Table 2). With the control for the general factor, the range of proportions of ECV of the specific factors was between 3.3 % and 14.6 % in parental reports and between 4.9 % and 17.6% in teacher reports. We recommend that the factors with lower than 4 % of ECV should not be used to calculate the score of a specific factor. Only the Behavioral Problems factor did not meet this requirement and only in parent reports. The criterion of 4 % is arbitrary, however we are not aware of any guideline to evaluate the proportion of ECV.

Discussion

Our analyses confirmed that among the classic five-factor, the three-factor, and a bifactor model, the bifactor structure of the SDQ yielded the closest fit to data, irrespective of the source of information. The bifactor model could account for the strong correlations usually observed among the more specific behavioral problems by including a common problem severity factor. Detailed inspection of factor loadings indicate that difficulties concerning hyperactive-impulsive behavior and conduct disorder are the most important or salient when a child's behavior is at evaluation. This pattern is more pronounced in the teacher model, and this finding is in line with previous research on the predictive power of teachers' expert knowledge to the impairment of executive functions. It seems that the impression of a teacher on the attention problems of a child – who learns and plays with peers in a structured school environment – is a good predictor of the "cool" executive functions (e.g., Kerr & Zelazo, 2004), and at the same time of the symptoms of ADHD (Jonsdottir, Bouma, Sergeant, & Scherder, 2006). Based on the results obtained by the bifactor model, an implication for screening and clinical assessment could be to regard the Total Difficulties Score more seriously at risk of various externalizing problems. On the basis of many studies (for details, see A. Goodman, et al., 2010) this score is associated with increasing rates of clinician-rated diagnoses of child mental disorder across its full range; however the score does not include the score of the Prosocial Scale which constitutes a part of our general factor. The obtained pattern also raises a problematic question about screening. It is probable that by applying this questionnaire, someone will find “at risk children” with a learning disorder or whose academic achievement is lower than the average without any special problematic domains. Nevertheless, the detection of extensive problems might be sufficient to get the children to obtain special education training. Concurrently, the bifactor structure proposes that subscales are worth reporting, as it strengthens the separation of the five symptom-clusters. According to scale score reliabilities, interpretation of specific subscales seems to be reasonable, and therefore we would recommend using both the global and specific factors in the diagnostic process. This practice becomes more pronounced in the presence and detection of problems concerning anxiety, depression, and peer relationships, especially in case of teachers as informants. The general factor explains much of the common variance in the manifest measures, and this phenomenon is slightly more pronounced in the parent model.

It is the first time that a bifactor model of SDQ has been estimated and compared to previously suggested measurement models. Although the classic first order five-factor structure demonstrated an acceptable fit to our data, the bifactor model yielded a significantly better fit. Our results definitely do not support the three-factor model.

The bifactor model of SDQ should be further replicated in larger samples and in clinic-referred samples. This questionnaire is primarily considered as a screening tool in community samples, but frequently used in clinical care to obtain supplementary information. When assessing the generalizability of the model an essential part of the work would be to test the effect of several forms of childhood developmental disorders on the model. Analysis of the “impact supplement” (items on overall distress and social impairment) might be a future task for the better understanding of symptom severity.

In the current study the estimation of probable developmental changes in the structure of strengths and difficulties was not possible. Thus, behavioral ratings of children from a wider age range must be collected. The application and validation of the self-report version of SDQ is emphasized in adolescent samples. Further research should clarify the gender invariance of the currently identified bifactor model. This is particularly important, because phenotypic forms of psychiatric syndromes could be distinct in boys and in girls. Similar steps should be taken to test the effect of socioeconomic status (e.g. Rothenberger, Becker, Erhart, Wille, & Ravens-Sieberer, 2008).

In summary, our study suggests that childhood behavior problems are best described by a bifactor model, which means that the measured construct has a dominant global factor and specific components that indicate multidimensionality. We consider the general factor as an indicator of problem or symptom severity, which refers to a global and primary belief about the child's functioning in his or her everyday environment. It is probable that this impression is mainly built up from frequent behaviors that happen in the presence of externalizing problems (Achenbach, et al., 2008). At the same time, the specific components cover the main childhood psychiatric disorders, which could manifest along a broader dimension, involving subclinical variants besides non-symptomatic cases and those with severe problems.

References

Achenbach, T. M., Becker, A., Dopfner, M., Heiervang, E., Roessner, V., Steinhausen, H. C., & Rothenberger, A. (2008). Multicultural assessment of child and adolescent psychopathology with ASEBA and SDQ instruments: research findings, applications, and future directions. *J Child Psychol Psychiatry, 49*(3), 251-275. doi: JCPP1867 [pii] 10.1111/j.1469-7610.2007.01867.x

Asparouhov, T., & Muthén, B. (2006). Robust Chi square difference testing with mean and variance adjusted test statistics. *Mplus Web Notes, 10*.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders, fourth edition, text revision (DSM-IV-TR)*. Arlington, VA: American Psychiatric Association.

Becker, A., Hagenberg, N., Roessner, V., Woerner, W., & Rothenberger, A. (2004). Evaluation of the self-reported SDQ in a clinical setting: do self-reports tell us more than ratings by adult informants? *Eur Child Adolesc Psychiatry, 13 Suppl 2*, II17-24. doi: 10.1007/s00787-004-2004-4

Bentler, P. (2009). Alpha, Dimension-Free, and Model-Based Internal Consistency Reliability. *Psychometrika, 74*(1), 137-143. doi: 10.1007/s11336-008-9100-1

Berge, J. M. F., & Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika, 69*(4), 613-625. doi: 10.1007/bf02289858

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Browne, M. V., & Cudek, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & S. J. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A Tutorial on Hierarchically Structured Constructs. *Journal of Personality, 80*(4), 796-846. doi: 10.1111/j.1467-6494.2011.00749.x

Cortina, J. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98-104. doi: citeulike-article-id:3800748

Dickey, W. C., & Blumberg, S. J. (2004). Revisiting the factor structure of the strengths and difficulties questionnaire: United States, 2001. *J Am Acad Child Adolesc Psychiatry, 43*(9), 1159-1167. doi: S0890-8567(09)61451-9 [pii]

10.1097/01.chi.0000132808.36708.a9

Finney, S. J., & DiStefano, C. (2006). Nonnormal and categorical data in structural equation modeling. In G. R. Hancock & R. D. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269-314). Greenwich, CT: Information Age.

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. *Ann Gen Psychiatry, 8*, 20. doi: 1744-859X-8-20 [pii]

10.1186/1744-859X-8-20

Gibbons, R. D., Bock, R. D., Hedeker, D., Weiss, D. J., Segawa, E., Bhaumik, D. K., . . . Stover, A. (2007). Full-Information Item Bifactor Analysis of Graded Response Data. *Applied Psychological Measurement, 31*(1), 4-19. doi: 10.1177/0146621606289485

Gibbons, R. D., & Hedeker, D. R. (1992). Full-Information Item Bi-Factor Analysis. *Psychometrika, 57*(3), 423-436.

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. *J Abnorm Child Psychol, 38*(8), 1179-1191. doi: 10.1007/s10802-010-9434-x

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol Psychiatry, 38*(5), 581-586.

Hagquist, C. (2007). The psychometric properties of the self-reported SDQ - An analysis of Swedish data based on the Rasch model. *Personality and Individual Differences, 43*(5), 1289-1301. doi: 10.1016/j.paid.2007.03.022

Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika, 2*, 41-54. doi: 10.1007/bf02287965

Hudziak, J. J., Achenbach, T. M., Althoff, R. R., & Pine, D. S. (2007). A dimensional approach to developmental psychopathology. *Int J Methods Psychiatr Res, 16 Suppl 1*, S16-23. doi: 10.1002/mpr.217

Jonsdottir, S., Bouma, A., Sergeant, J. A., & Scherder, E. J. A. (2006). Relationships between neuropsychological measures of executive function and behavioral measures of ADHD symptoms and comorbid behavior. *Archives of Clinical Neuropsychology, 21*(5), 383-394. doi: 10.1016/j.acn.2006.05.003

Kerr, A., & Zelazo, P. D. (2004). Development of 'hot' executive function: The children's gambling task. *Brain and Cognition, 55*(1), 148-157. doi: 10.1016/s0278-2626(03)00275-6

Koskelainen, M., Sourander, A., & Vauras, M. (2001). Self-reported strengths and difficulties in a community sample of Finnish adolescents. *Eur Child Adolesc Psychiatry, 10*(3), 180-185.

Krueger, R. F., McGue, M., & Iacono, W. G. (2001). The higher-order structure of common DSM mental disorders: internalization, externalization, and their connections to personality. *Personality and Individual Differences, 30*(7), 1245-1259. doi: 10.1016/s0191-8869(00)00106-9

Martel, M. M., Gremillion, M., Roberts, B., von Eye, A., & Nigg, J. T. (2010). The structure of childhood disruptive behaviors. *Psychol Assess, 22*(4), 816-826. doi: 2010-24850-003 [pii] 10.1037/a0020975

Nigg, J. T., Tannock, R., & Rohde, L. A. (2010). What Is to Be the Fate of ADHD Subtypes? An Introduction to the Special Section on Research on the ADHD Subtypes and

Implications for the DSM–V. *Journal of Clinical Child & Adolescent Psychology, 39*(6), 723-725. doi: 10.1080/15374416.2010.517171

Paulhus, D. L., Robins, R. W., Trzesniewski, K. H., & Tracy, J. L. (2004). Two Replicable Suppressor Situations in Personality Research. *Multivariate Behavioral Research, 39*(2), 303 - 328.

Reise, S. P. (in press). The Rebirth of Bifactor Measurement Models. *Multivariate Behavioral Research*.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor Models and Rotations: Exploring the Extent to Which Multidimensional Data Yield Univocal Scale Scores. *Journal of Personality Assessment, 92*(6), 544-559. doi: 10.1080/00223891.2010.496477

Reise, S. P., Morizot, J., & Hays, R. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research, 16*(0), 19-31. doi: 10.1007/s11136-007-9183-7

Rhee, S. H., Willcutt, E. G., Hartman, C. A., Pennington, B. F., & DeFries, J. C. (2008). Test of Alternative Hypotheses Explaining the Comorbidity between Attention-Deficit/Hyperactivity Disorder and Conduct Disorder. *Journal of Abnormal Child Psychology, 36*(1), 29-40.

Riso, D. D., 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*(6), 570-575. doi: 10.1016/j.paid.2010.05.005

Ronning, J. A., Handegaard, B. H., Sourander, A., & Morch, W. T. (2004). The Strengths and Difficulties Self-Report Questionnaire as a screening instrument in Norwegian community samples. *Eur Child Adolesc Psychiatry, 13*(2), 73-82. doi: 10.1007/s00787-004-0356-4

Rothenberger, A., Becker, A., Erhart, M., Wille, N., & Ravens-Sieberer, U. (2008). Psychometric properties of the Parent Strengths and Difficulties Questionnaire in the general population of German children and adolescents: Results of the BELLA study. *European Child & Adolescent Psychiatry, 17*(Suppl 1), 99-105. doi: 10.1007/s00787-008-1011-2

Ruchkin, V., Koposov, R., & Schwab-Stone, M. (2007). The Strength and Difficulties Questionnaire: scale validation with Russian adolescents. *J Clin Psychol, 63*(9), 861-869. doi: 10.1002/jclp.20401

Sinharay, S., Haberman, S., & Puhan, G. (2007). Subscores Based on Classical Test Theory: To Report or Not to Report. *Educational Measurement: Issues and Practice, 26*(4), 21-28. doi: 10.1111/j.1745-3992.2007.00105.x

Stansfeld, S. A., Clark, C., Caldwell, T., Rodgers, B., & Power, C. (2008). Psychosocial work characteristics and anxiety and depressive disorders in midlife: the effects of prior psychological distress. *Occupational and Environmental Medicine, 65*(9), 634-642. doi: 10.1136/oem.2007.036640

Thomas, M. L. (2012). Rewards of bridging the divide between measurement and clinical theory: Demonstration of a bifactor model for the Brief Symptom Inventory. *Psychological Assessment, 24*(1), 101-113. doi: 10.1037/a0024712

Valo, S., & Tannock, R. (2010). Diagnostic Instability of DSM–IV ADHD Subtypes: Effects of Informant Source, Instrumentation, and Methods for Combining Symptom Reports. *Journal of Clinical Child & Adolescent Psychology, 39*(6), 749-760. doi: 10.1080/15374416.2010.517172

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. *J Child Psychol Psychiatry, 49*(12), 1304-1312. doi: JCPP1942 [pii] 10.1111/j.1469-7610.2008.01942.x

Warnick, E. M., Bracken, M. B., & Kasl, S. (2008). Screening Efficiency of the Child Behavior Checklist and Strengths and Difficulties Questionnaire: A Systematic Review. *Child and Adolescent Mental Health, 13*(3), 140-147. doi: 10.1111/j.1475-3588.2007.00461.x

Yung, Y.-F., Thissen, D., & McLeod, L. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika, 64*(2), 113-128. doi: 10.1007/bf02294531