Running head: French SDQ – Teacher version

Investigation of a bifactor model of the Strengths and Difficulties Questionnaire

Hervé CACI[1*], Alexandre J. S. MORIN[2], & Antoine TRAN[1]

[*]Correspondance to :
[1]Hervé CACI, MD, PhD
Hôpitaux Pédiatrique de Nice – CHU Lenval
57, avenue de la Californie
06200 Nice
Tel : +33 (0)4 92 03 08 71
Fax : +33 (0)4 92 03 05 58
Email : caci.h@pediatrie-chulenval-nice.fr

[2]Institute for Positive Psychology and Education
Australian Catholic University
Australia

Acknowledgements

The ChiP-ARD study was founded by a grant to HC from the French Health Ministry and is recorded on clinicaltrials.gov under the reference NCT01260792.The authors are grateful to Eric FONTAS, MD for his help in setting up the study procedures, to Vanina OLIVERI and Kevin DOLLET for their help in the data collection and coordination, to the *Inspection Académique des Alpes-Maritimes* and to the *Rectorat des Alpes-Maritimes et du Var* for their valuable support, and to the teachers, pupils and parents for participating in this study.

This is the final prepublication version of :

Caci, H., **Morin, A.J.S.,** & Tran, A. (2015). Investigation of a bifactor model of the Strengths and Difficulties Questionnaire. *European Child and Adolescent Psychiatry*, *24* (10), *1291-1301* DOI 10.1007/s00787-015-0679-3

Abstract

Background. The Strengths and Difficulties Questionnaire (SDQ) is used to measure psychopathological symptoms in children and adolescents from 4 to 17 years old, but its underlying structure is still a matter of debate. Indeed, on the basis of a systematic review of English and non-English articles conducted using multiple databases, 54 studies reporting on the factor structure of the SDQ were located. The original 5 first-order factor structure is generally supported by exploratory procedures, but support based on confirmatory factor analyses is not as clear. **Method.** We analysed data from 889 youths from the general French population, rated on the SDQ by their teachers. We tested the original model, hierarchical models and bifactor models. **Results.** The best-fitting model is a bifactor model with the five *a priori* factors grouped in two Global-factors (Externalizing Disorders — Hyperactivity and Conduct — and Internalizing Disorders — Peer relationships and Emotions) and one Strength/Prosocial factor. However, we show that the Conduct Specific-factor should not be used in practice in its current state, that the Hyperactivity Specific-factor mainly covers hyperactivity rather than inattention, and that the Peer Problems Specific-factor mainly reflects a preference for solitude. Nevertheless, the measurement model proved to be fully invariant across gender and school levels (kindergarten, primary and secondary schools), with statistically significant differences in latent means between genders only. **Conclusions.** Beyond computing the five *a priori* scores when using the teacher-ratings of the SDQ, our results prove the usefulness of computing Externalizing Disorders and Internalizing Disorders global scores.

Keywords:

Structural Equation Modeling, Teachers, Externalizing Disorders, Internalizing Disorders, Prosocial Behaviour, SDQ, Measurement Invariance

Typically, behavioral disorders are classified into two broad categories: Externalizing Disorders (ED - such as Attention-Deficit Hyperactivity Disorder - ADHD, Conduct Disorders - CD, and Oppositional-Defiant Disorder - ODD) and Internalizing Disorders (ID - such as Anxiety Disorders and Major Depression)[1]. Although structured diagnostic interviews are clearly the gold standard for assessing these disorders [2], these procedures are costly. Thus, self- or informant- reported questionnaires are often used as a first step in community screenings or large-scale community studies [3].

The Strengths and Difficulties Questionnaire (SDQ; [4] was developed as an extension of Rutter's parent questionnaire [5,6], and has become one of the most commonly used instrument for measuring psychopathological symptoms in school-age children and adolescents. All versions of the SDQ (parents and teacher for children and adolescents aged between 4 and 17 years, and self-report for adolescents aged between 11 and 17 years) count 25 items rated on a 3-point scale ("Not true", "Somewhat true", or "Certainly true"). Five items (7, 11, 14, 21, 25) are reversed scored. The SDQ is available free of charge for non-commercial purposes (www.sdqinfo.com) in 40 languages [7]. Based on initial principal component analyses (PCA), five component scores are generally formed. Four reflect behavioural Difficulties (Emotional Symptoms, Conduct Problems, Hyperactivity-Inattention and Peer Problems) and one reflects behavioural Strengths (assessed through Prosocial behaviours). It logically follows that the former four subscales should combine into a higher-order Difficulty factor negatively correlated with the latter Strengths/Prosocial Behaviours (S/PB) factor. Another possible structure would be to combine the Hyperactivity-Inattention and Conduct Problem subscales into a higher-order ED factor, and the Emotional Symptoms and Peer Problems subscales into an higher-order ID factor [8]. Although these alternative structures make sense and each have received some support, there appears to be a need for clarification regarding the optimal factor structure of the SDQ.

For this purpose, we systematically searched MEDLINE, EMBASE, ERIC, PSYCInfo and ScienceDirect databases for papers published between January 1st 1995 and December 31st 2013, using the strings ["SDQ" or "Strengths and Difficulties questionnaire"] and ["factor analysis" or "factor structure"]. Based on available abstract and/or full-text content, we discarded papers: (a) reporting on a different instrument with the same acronym (e.g., the Self-Description Questionnaire), (b) with a main focus that was not the structure of the SDQ; (c) that did not analyse all 25 items of the SDQ; (d) only reported research conducted on special populations (e.g. intellectually disabled children) or preschool children using a specific version of the SDQ. This systematic review encompasses 54 publications. These publications are summarized in Tables S1 (self-report version), S2 (parent version), and S3 (teacher version) of the online supplements, together with their references. For each

study, we report the sample sizes, age range, language version, country where the study was conducted, and whether analyses were conducted in specific subsamples (e.g. males and females). We also report on the method used to analyse the data (exploratory factor analyses, principal component analyses, confirmatory factor analyses), and whether the number of factor was fixed a priori to 5 in a confirmatory manner, or whether an exploratory approach was used to determine the number of factors.

First-Order Factor Structure

Studies provide some support to the *a priori* first-order 5-factor structure but many suffer from important limitations, such as the reliance on Principal Component Analyses (PCA) which are not suited to the analysis of the underlying structure of psychological constructs [9] and the reliance on exploratory procedures when an *a priori* structure has been previously defined [10,11]. Furthermore, studies relying on Confirmatory Factor Analyses (CFA) failed to provide a clear and unmitigated support to the adequacy of this *a priori* structure, showing the fit to be "good" in 20 cases, "acceptable" in 22 cases and "poor" in 24 cases.

Studies also generally failed to support the adequacy of two alternative 3-factor structures including 3 factors reflecting S/PB, ID and ED: one where S/PB is defined using the five *a priori* items (1, 4, 9, 17, 20), and the Dickey and Blumberg (2004) model where the S/PB was refined through exploratory procedures and counted 8 items (adding items 7, 11, 14) [12]. In model the ED factor was defined through 9 items (2, 5, 10, 12, 15, 18, 21, 25 and item 7 which cross-loaded on S/PB), the ID factor was defined through 8 items (3, 6, 8, 13, 16, 19, 23 and 24), and item 22 ("Steals from home, school or elsewhere") was discarded.

Higher-Order Factor Structure

A total of 6 studies estimated a model including one higher-order Difficulties factor, estimated from the four first-order factors, and allowed to correlate with the S/PB factor. Only one of these studies clearly supported this model, whereas three clearly failed to support this model. Another model including two correlated higher-order factors representing ED and ID, allowed to correlate with the S/PB factor, was tested in 3 studies. These studies only provided partial support to this structure limited to specific subsamples or SDQ versions. Thus, although the question of whether SDQ items form global constructs over and above the five specific subscales appears important, a question that remains is whether a higher-order model is the best way to explore this issue.

Alternative Representations

In psychiatric measurement, a crucial question is whether a primary dimension (e.g., ED) exists as a unitary construct including specificities, or whether these specificities rather define distinct facets without a common core (i.e., a first-order CFA). Higher-order models, where higher-order factors are defined from the

covariance among first-order factors, represents one way of looking at this issue [13,14]. However, bifactor models provide a more flexible alternative [15,16] based on the assumption that a *f*-factor solution exists for a set of *n* items with one global (G) factor and *f*-1 specific (S) factors. Bifactor models can easily be expanded to include more than one G-factor. The S-factors are typically specified as uncorrelated (orthogonal) to one another and with the G-factor(s). The Schmid-Leiman transformation (SLT) [17] can be used to convert a higher-order model to a bifactor approximation. However, each item's association with the SLT G- and S- factors are obtained by multiplying their first-order loadings by constants, resulting in a ratio of G to S factor loadings that is exactly the same for all items associated with a first-order factor. This is a reason why true bifactor models are more flexible, and tend to provide a better fit to the data, than higher-order models [13,15,18]. Recently, bifactor models have been found to provide superior representation of ADHD [19,20] and depression [21] than higher-order factors. Similar results have also been found for the SDQ, unfortunately without using all 25 items [22].

Gender and Age/Grade Similarities and Differences

An important test of the generalizability of a measurement model has to do with the possibility to replicate results across multiple meaningful subgroups of participants and the demonstration that meaningful unbiased group comparisons are possible. This verification requires systematic tests of measurement invariance [23,24]. Among the reviewed studies, a handful separately estimated, and compared, the SDQ measurement models across meaningful subgroups of participants (origin, gender, age or grade, combinations) (Niclasen *et al*., 2013, Niclasen *et al.*, 2012). These studies generally report similar measurement models across subgroups, although they sometimes suggest variations as a function of age or grade. However, the results from the 9 studies that conducted systematic tests of measurement invariance generally supported some level of invariance of the SDQ measurement model across genders, age/grade, language, or informant.

An important test of the discriminant validity of a measure lies in its ability to recover group differences in the constructs of interest. Interestingly, measurement invariance should be verified prior to tests of group-based mean-difference. In relations to the SDQ, English norms (http://www.sdqinfo.com/norms/UKNorm2.pdf) show that boys tend to present higher levels than girls on the first-order Conduct (Cohen's d=.39), Hyperactivity (Cohen's d=.60), and Peer Relationships (Cohen's d=.17) factors, and lower levels on the S/PB (Cohen's d=.56) factor. Similarly, younger children are known to present higher levels on the Emotional (Cohen's d=.11), Hyperactivity (Cohen's d=.15) and S/PB (Cohen's d=.08) first-order factors, albeit the effect sizes are negligible (http://www.sdqinfo.com/norms/UKNorm3.pdf). Among the reviewed studies, only one systematically explored latent means differences as a function of gender, after having established the measurement invariance of the

model [25]. This study showed that boys tended to present higher scores on the Conduct Problems, Hyperactivity-Inattention, and Peer Problems factors than girls, who tended to present higher levels of S/PB.

The present study

In the present study, we aim to provide a comprehensive test of the complete factor structure of the teacher version of the French SDQ. After contrasting alternative representations of the first-order structure, we investigate the more global constructs present in the SDQ (i.e., ED, ID or Difficulties) using alternative higher-order and bifactor models. We then test whether the best-fitting model is invariant across groups formed on the basis of gender (boys/girls) and school level (kindergarten, primary, secondary) to ascertain whether answers provided to the SDQ can be meaningfully compared across these groups. We then verify whether well-documented group-based differences, or lack thereof, in latent means can be replicated.

Methods

Participants, Material, and Procedures

This paper uses data from the ChiP-ARD (*Children and Parents with ADHD and Related Disorders*) study, targeting French children and adolescents from the general population aged between 4 to 18 years old [19,20,26]. Overall, 262 teachers participated in the study (mean age=43.9; S.D.=8.6; range=24-61); forty-seven were males (17.94%). Each was asked to rate 2 to 4 youths from their classes whose name began with a letter randomly drawn from the alphabet. The official French adaptation of the teacher version of the SDQ for 4- to 17-year olds was obtained from the official website (http://www.sdqinfo.org). SDQ ratings were returned for a total of 889 youths (including 455 girls, 51.18%): 132 attended kindergarten (14.85%; including 64 girls), 350 attended primary schools (39.37%; including 174 girls), and 407 attended secondary schools (45.78%; including 217 girls). Girls were aged on average 5.69 (SD = .29) years in kindergarten, 8.62 (SD = 1.54) years in primary school, and 13.87 (SD = 2.16) in secondary school. Boys were aged on average 5.65 (SD = .36) in kindergarten, 8.61 (SD = 1.51) years in primary school, and 13.47 (1.75) years in secondary school. The Commissioner of Education and the Department of Education supported this study that complied with normative ethical prescriptions for French medical research. The *Commission Nationale Informatique et Liberté* approved the procedures used to keep the data secured and anonymous.

Analyses

The main models were estimated with Mplus 7.11 [27] from polychoric correlation matrices using robust weight least square (WLSMV) estimation, which has been found to outperform Maximum Likelihood with ordered-categorical items involving 5 or less answers categories [28-32]. The fit of 4 alternative *a priori* first-

order models was contrasted: (M1) a one-factor model defined based on all SDQ items; (M2) a model including 2 correlated factors (Strengths, defined based on all 5 prosocial items, and Difficulties, defined based on the other 20 items); (M3) a model including 3 correlated factors [ID (10 items), ED (10 items), and S/PB (5 items)]; (M4) a 3-factor model defined according to Dickey and Blumberg (2004) specifications [ID (8 items), ED (9 items), and S/PB (8 items)]; (M5) the *a priori* SDQ model including 5 correlated factors (5 items each). Assuming that the *a priori* 5-factor model provides the highest level of fit to the data, two higher-order factor models will be contrasted: (M6) a model including a single higher-order factor defined on the basis of the 4 first-order Difficulty factors (Emotional Symptoms, Conduct Problems, Hyperactivity-Inattention and Peer Problems) and correlated to a S/PB factor; (M7) a model including two correlated higher-order Internalizing (defined on the basis of the Emotional Symptoms and Peer Problems first-order factors) and Externalizing (defined on the basis of the Conduct Problems and Hyperactivity-Inattention first-order factors) Disorders factors, correlated to a S/PB factor. Likewise, two bifactor models will be contrasted: (M8) a model including a single Difficulty G-factor defined on the basis of all items associated with four S-factors and correlated to a S/PB factor; (M9) a model including two correlated ID (defined on the basis of all items associated with the Emotional Symptoms and Peer Problems S-factors) and ED (defined on the basis of all items associated with the Conduct Problems and Hyperactivity-Inattention S-factors) G-factors, correlated to a S/PB factor.

From the best-fitting model, we will then perform tests of measurement invariance tests across gender (male versus females) and school level (kindergarten, primary school, secondary school) following Meredith recommendations [23] as adapted for ordered-categorical items [21,33]. The sequence of tests is as follows: (i) configural invariance, (ii) metric/weak invariance (invariance of the factor loadings); (iii) scalar/strong invariance (invariance of the factor loadings and thresholds); (iv) strict invariance (invariance of the factor loadings, thresholds and uniquenesses), (v) invariance of the latent variances-covariances (invariance of the factor loadings, thresholds, uniquenesses and variances-covariances), and (vi) latent means invariance (invariance of the factor loadings, thresholds, uniquenesses, variances and latent means).

The fit of all models was evaluated using the WLSMV Chi-square statistic (χ^2), the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), the Root Mean Square Error of Approximation (RMSEA) and its 90% confidence interval [34,35]. Values greater than .95 for CFI and TLI are considered to be indicative of adequate model fit. Values smaller than .08 or .06 for the RMSEA support respectively acceptable and excellent model fit. Chi-square differences tests were computed using MPlus DIFFTEST function (MD$\Delta\chi^2$; [36,37], with the significance level to identify non-invariance fixed at .01 to take into account the overall number of MD$\Delta\chi^2$ tests

performed [38-40]. Because the χ^2 and MD$\Delta\chi^2$ are oversensitive to sample size and to minor model misspecifications, additional indices were used in the comparisons of nested invariance models. Thus, a CFI diminution of .01 or less and a RMSEA augmentation of .015 or less between a model and the preceding model in the invariance sequence indicate that the invariance hypothesis should not be rejected [41,42].

Scale score reliability for the estimated factor are estimated based on omega (ω) [43] which has the advantage over traditional scale score reliability estimates (e.g. Cronbach's α) to take into account the strength of association between items and all constructs, as well as item-specific measurement errors [44]. This makes it more realistic for complex measurement models such as those considered here.

Results

The results reported in Table 1 show that, among the alternative first-order models, only M5 reaches an acceptable, albeit marginal, fit to the data. The parameter estimates from this model are reported in Table 2. All items presented high and satisfactory factor loadings on their main factors (λ=0.413 to 0.951; M=0.746) except item 23 (λ=0.248). All scale score reliability coefficients proved satisfactory (ω=.758 to .914). However, the correlations between the Conduct Problems and Hyperactivity-Inattention factors (r=.748) and between the Emotional Symptoms and Peer Problems (r=.513) factors, as well as the marginal fit of the model, suggest a need for further verifications. More precisely, these correlations are in line with the idea that these two pairs of factors may in fact assess two overarching constructs of, respectively, ED and ID. Although some of the other correlations (for instance those involving the S/PB factor) are of a similar magnitude, there are no a priori theoretical reasons to expect that these other factors would also form a reduced set of overarching constructs.

Among the two alternative higher-order models, M6 failed to reach an acceptable level of fit to the data, whereas M7 provided an acceptable, yet marginal, fit to the data that could not be empirically distinguished from M5 (ΔCFI=-.002; ΔTLI=-.002; ΔRMSEA=+.001). Similarly, bifactor model M8 provided a marginal level of fit to the data that could not be empirically distinguished from either M5 or M7. In contrast, the fit of bifactor model M9 proved fully acceptable and substantially better than the fit of alternative models M5, M7 and M8 according to the goodness-of-fit indices (ΔCFI=+.018 to +.023; ΔTLI=+.019 to +.021; ΔRMSEA=-.007 to -.008), although the confidence interval of the RMSEA suggest that these differences may not be fully significant.

Model M9 was thus retained (Figure I), and corresponding parameter estimates reported in Table 3 show that the S/PB factor and the ED G-factor are well-defined through relatively high factor loadings (λ=.679 to .830 and λ=.467 to .791, respectively), and present a satisfactorily high level of scale score reliability (ω=.883 and ω=.926, respectively). Noticeably, over and above their associations with this ED G-factor, most of these items

present a relatively low level of specificity associated with the Conduct Problems (λ=-.119 to .492) and Hyperactivity-Inattention (λ=-.161 to .758) S-factors, themselves defined mostly by relatively low loadings. The two items with the greatest level of specific association with the Conduct Problems S-factor concern covert – and thus less Exteriorized – forms of violence (*often lies or cheats*, λ=.492; *steals from home, school or elsewhere*, λ=.439). As a result, the scale score reliability of this S-factor remains quite low (ω=.544). Similarly, two items present a high level of specificity on the Hyperactivity-Inattention S-factor, and both assess symptoms of Hyperactivity (*restless, overactive, cannot stay still for long*, λ=.622; *constantly fidgeting or squirming*, λ=.758), rather than Inattention. Given the magnitude of these two specific loadings, the scale score reliability of this S-factor remains generally satisfactory (ω=.705).

The ID G-factor is not as well defined as the ED G-factor, although it still reflects reasonably well a common core of ID manifestations. This G-factor is defined though: (a) 4 items with λ>.500 covering manifestations of social rejection (*has at least one good friend*; *generally liked by other children*; *picked on or bullied by other children*) and generic unhappiness (*often unhappy, depressed or tearful*); (b) 4 items with .200≤λ<.500 covering manifestations of anxiety (*often complains of headaches, stomach-aches or sickness; nervous in new situations, easily loses confidence*; *many fears, easily scared*) and preference for solitude (*would rather be alone than with other youth*); (c) Two items with low or non-significant factor loadings covering a generic tendency to worry (*many worries or often seems worried*) and preference for adult company (*gets along better with adults than with other children*). Supporting the potential usefulness of this G-factor, its model-based scale-score reliability appears fully satisfactory (ω=.804). The S-factor reflecting Emotional Symptoms going over and above this generic presence of ID is also well-defined through high factor loadings (λ=.468 to .780), and a satisfactory level of scale score reliability (ω=.832). In contrast, the Peer Problems S-factor appears to be more strongly defined through items reflecting a preference for solitude from peers (*would rather be alone than with other youth*, λ=.570; *gets along better with adults than with other children*, λ=.707) than items reflecting peer rejection (λ=.123 to .313) which present stronger relations with the ID G-factor – resulting in a lower S-factor scale score reliability estimate (ω=.650).

Results for model M9 (Table 1) supported the complete invariance of the measurement model, as well as the invariance of the variances and covariances across gender groups and for all school levels considered (kindergarten, primary, secondary); some MD$\Delta\chi$2 proved significant but none of the ΔCFI, ΔTLI and ΔRMSEA exceeded the recommended cut-offs. The teachers' version of the SDQ thus provides results that are fully comparable across male and female youths attending kindergarten, primary schools, and secondary schools. The

results also support the absence of latent means differences across school levels (ΔCFI=.000; ΔTLI=+.001; ΔRMSEA=-.001), and the presence of latent means differences across genders (ΔCFI=-.017; ΔTLI=-.016; ΔRMSEA=+.007). When girls latent means are fixed to 0 for identification purposes and differences are expressed in standard deviation units, boys have higher latent means on the ED G-factor (0.493, $p<.05$), the ID G-factor (0.252, $p<.05$) and on the Hyperactivity/Inattention S-factor (0.287, $p\leq.05$), but lower latent means on the Strength/Prosocial Behaviour factor (-0.467, $p<.05$) and the Emotional Problems S-factor (-0.231, $p<.05$). Latent means did not differ across genders on the Conduct Problems and Peer Problems S-factors.

DISCUSSION

The factor structure of all SDQ versions has been extensively cross culturally assessed. Our review showed that the *a priori* first-order 5-factor model has generally received strong support. In contrast, research results are mixed regarding the presence of more global constructs reflecting ID, ED or global Difficulties [8]. This could partly be related to the reliance on higher-order factor models over bifactor models [13,15,18].

In the present study, we explored the global and specific factor structure of the SDQ using data from the general population. Alternative first-order CFA models were first contrasted to verify the adequacy of the *a priori* 5-factor model over alternative models. Examination of the parameter estimates revealed well-defined factors and satisfactory model-based estimates of scale-score reliability. However, the fit of this model remained close to the lowest bound of acceptability according to conventional guidelines. Furthermore, the estimated factor correlations suggested exploring the presence of more global constructs.

None of the higher-order representations of the SDQ considered provided a satisfactory alternative to the *a priori* first-order factor model. Conversely, a bifactor model including two correlated G-factors reflecting ID and ED (themselves correlated to a S/PB first-order factor) provided a satisfactory level of fit to the data and a clear improvement over the fit of the first-order factor model. It should be noted that this conclusion is limited by the fact that the confidence intervals for the RMSEA mainly overlapped across models – although the efficacy of this specific indicator of model fit has yet to be systematically investigated in the context of WLSMV estimation. Similarly, some of the estimated factor loadings for this model turned out to be non-significant, which is consistent with the nature of bifactor models where each item cannot realistically be assumed to present equally strong associations with Global and Specific factors (Morin, Arens, & Marsh, 2015). Rather, the specific patterns of significant versus non-significant loadings helped us to refine the interpretation of the G- and S-factors.

Parameters estimates from the final model revealed three well-defined S/PB, ED and ID factors, although the ID G-factor mainly reflects the social rejection and anxiety components of ID. These three factors also

present high and satisfactory scale score reliability (ω=.804 to .926), supporting the use of the corresponding total score in research and practice. Similarly, the S-factor reflecting Emotional Symptoms is well-defined through high factor loadings from all of items, and presents satisfactory scale score reliability (ω=.832), which confirms the importance of using scores on this factor to complement ID ratings obtained based on the G-factor.

The S-factors reflecting the *a priori* SDQ scales do not appear to be defined as well as the G-factors, but still convey meaningful specificity over and above the assessment provided by the G-factors. Whereas the G-factor appears to provide a relatively complete overarching assessment of ED, the S-factor related to Conduct Problems mainly reflects covert forms of conduct disorders related to stealing and cheating going beyond the more overt forms of violence that are specifically covered by the ED G-factor. Similar distinctions between overt and covert manifestations of violence are often noted in the research literature on Conduct Disorders [45-47]. However, although this distinction appears worthy of consideration in the measurement model of the SDQ as a way to control for the specificity of these covert behaviours beyond what is already assessed through the ED G-factor, the low scale score reliability of this S-factor (ω=.544) suggests that this specific Conduct Problem subscale should not be used in practice in its current state. Rather, future research should seek ways to improve the assessment of covert behaviours in order increase the meaningfulness of this subscale. One hypothesis may be that teachers alone cannot capture all facets of overt/covert behaviours, and that one way of improving this assessment might be to use multiple informants across different settings. This would be expected based on multi-rater studies of conduct problems and antisocial behaviours [48,49].

The Hyperactivity-Inattention S-factor mostly covers manifestations of Hyperactivity, rather than Inattention, going beyond the common core of ED. This is in line with the subtype-specificity of ADHD proposed by the DSM-5 [1], stating that ADHD symptoms can be dominated either by Hyperactivity or Inattention, as well as with previous bifactor representations of ADHD [19,20]. Similarly, the Peer Problems S-factor is mainly defined through items reflecting a preference for solitude, rather than the social rejection component of peer-related problems covered within the ID G-factor. The distinction between peer rejection and preference for solitude has been found to have important substantive implications in previous research on peer problems [50,51]. Although the results from the first-order 5-factor model suggested a need for the re-assessment of item 23 (“Gets on better with adults than with other children”) due to a very low factor loading on its *a priori* factor, the results from the final retained bifactor model rather suggest that this item plays an important role in the definition of this S-factor reflecting a preference for solitude (λ=.707). The level of specificity related to these S-factors (Hyperactivity-Inattention and Peer Problems) are sufficient to provide satisfactory scale score

reliability estimates that fully justify their use to complement ratings on the G-factors, in order to assess hyperactivity (ω=.705) and preference for solitude (ω=.650), over and above levels of ED and ID.

As a preliminary test of generalizability, we conducted tests of measurement invariance of the obtained factor structure across subgroups of participants. This verification is particularly relevant for the SDQ, which has been developed to be suitable for youth aged between the ages of 4 and 17, thus relying on the assumption that SDQ ratings would be comparable across this full developmental period. In line with these expectations, the final retained bifactor model proved to be fully invariant (configuration, loadings, thresholds, uniquenesses, and even variances and covariances) across genders and the three school levels considered. We further verified whether the latent means obtained on the estimated factors would replicate the results from previous studies as a test of the discriminant validity of the model. Interestingly, supporting previous studies, no mean-levels differences could be observed as a function of school levels. Conversely, but also supporting previous studies based either on the SDQ [25] or other instruments [19,20], our results showed that boys had higher levels of ED, ID, and Hyperactivity-Inattention, while they had lower levels of S/PB and Emotional Problems than girls.

In summary, our study shows that it is legitimate to compute five *a priori* scores when using teacher-ratings of the SDQ. Additionally, it may be even more informative to compute scores of ED and ID and then to interpret subscale-specific scores on the Conduct Problems, Hyperactivity-Inattention, Peer Problems and Emotional Symptoms factors as a function of the information they add to refine initial interpretations based on the ED and ID scores. More precisely, our results suggest that the Emotional Symptoms score would be meaningful in its own right as the content of this subscale is only imperfectly reflected in the ID factor. Conversely, the Conduct Problems, Hyperactivity-Inattention, Peer Problems apparently mainly, and respectively, reflect Covert Behaviours, Hyperactivity, and Preference for Solitude once global scores on the ED and ID factors are taken into account. Although our results are promising, future studies should still investigate the validity, sensibility, and specificity of SDQ assessments based on teachers-, parents- and self- ratings of the same instrument, and formal clinical assessments conducted using structured interviews. Indeed, although this study focused on the psychometric properties of the teacher's version of the SDQ, in practice, the assessment behavioural disorders typically seeks to identify behaviours that are pervasive across settings and thus aims to integrate multiple sources of information (parent, teachers, clinicians, self). This is important as these informants are known to provide different perspectives on the behaviours being rated (Merwood et al., 2013), due to their reliance on distinct frames of references in their interaction with the child being rated. In addition, the generalizability of our results to other versions of SDQ and other linguistic groups should also be investigated.

References

1. American Psychiatric Association (2013) Diagnostic and Statistical Manual of Mental Disorders, 5th edition. American Psychiatric Press, Washington,DC

2. Preisig M, Waeber G, Vollenweider P, Bovet P, Rothen S, Vandeleur C, Guex P, Middleton L, Waterworth D, Mooser V, Tozzi F, Muglia P (2009) The PsyCoLaus study: methodology and characteristics of the smaple of a population-based survey on psychiatric disorders and their association with genetic and cardiovascular risk factors. BMC Psychiatry 9 (9)

3. Kessler RC, Adler L, Ames M, Demler O, Faraone S, Hiripi E, Howes MJ, Jin R, Secnik K, Spencer T, Ustun TB, Walters EE (2005) The World Health Organization adult ADHD self-report scale (ASRS): a short screening scale for use in the general population. Psychological Medicine 35 (2):245-256

4. Goodman R (1997) The Strengths and Difficulties Questionnaire: a research note. Journal of Child Psychology and Psychiatry 38 (5):581-586

5. Rutter M (1967) A children's behaviour questionnaire for completion by teachers: preliminary findings. Journal of Child Psychology and Psychiatry 8 (1):1-11

6. Goodman R (1994) A modified version of the Rutter Parent Questionnaire including extra items on children's strengths: a research note. Journal of Child Psychology and Psychiatry 35 (8):1483-1494

7. Goodman R (2001) Psychometric properties of the strengths and difficulties questionnaire. Journal of the American Academy of Child and Adolescent Psychiatry 40 (11):1337-1345

8. Goodman A, Lamping DL, Ploubidis GB (2010) When to use broader internalising and externalising subscales instead of the hypothesized five subscales on the Strengths and Difficulties Questionnaire (SDQ): data from British parents, teachers and children. Journal of Abnormal Child Psychology 38 (8):1179-1191

9. Widaman KF (2007) Common factors versus components: principals and principles, errors and misconceptions. In: Cudeck R, MacCallum RC (eds) Factor analysis at 100. Lawrence Erlbaum Associates, Mahwah,NJ, pp 177-203

10. Morin AJS, Marsh HW, Nagengast B (2013) Exploratory Structural Equation Modeling. In: Hancock GR, Mueller RO (eds) Structural Equation Modeling: A Second Course. 2nd edn. IAP, Greewich,Connecticut, pp 395-436

11. Guay F, Morin AJS, Litalien D, Valois P, Vallerand RJ (In Press) An application of exploratory structural equation modeling to evaluate the construct validity and the gender invariance of the Academic Motivation Scale. The Journal of Experimental Education

12. Dickey WC, Blumberg SJ (2004) Revisiting the factor structure of the Strengths and Difficulties Questionnaire: United States, 2001. Journal of the American Academy of Child and Adolescent Psychiatry 43 (9):1159-1167

13. Brunner M, Nagy G, Wilhelm O (2012) A tutorial on hierarchically structured constructs. Journal of Personality 80 (4):796-846. doi:doi: 10.1111/j.1467-6494.2011.00749.x

14. Rindskopf D, Rose T (1988) Some theory and application of confirmatory second-order factor analyses. Multivariate Behavioral Research 23 (1):51-67

15. Chen FF, West SG, Sousa KH (2006) A comparison of bifactor and second-order models of quality of life. Multivariate Behavioral Research 41 (2):189-225

16. Reise SP, Moore TM, Haviland MG (2010) Bifactor models and rotations: exploring the extent to which multidimensional data yield univocal scale scores. Journal of Personality Assessment 92 (6):544-559

17. Schmid J, Leiman JM (1957) The development of hierarchical factor solutions. Psychometrika 22 (1):53-61

18. Reise SP (2012) Invited paper: The rediscovery of bifactor measurement models. Multivariate Behavioral Research 47 (5):667-696

19. Caci H, Morin AJS, Tran A (In press) Teacher ratings of the ADHD-RS IV in a community sample: Results from the ChiP-ARD study. Journal of Attention Disorders

20. Morin AJS, Tran A, Caci H (In press) Factorial Validity of the ADHD Adult Symptom Rating Scale in a French community sample: Results from the ChiP-ARD study. Journal of Attention Disorders

21. Morin AJS, Moullec G, Maïano C, Layet L, Just J-L, Ninot G (2011) Psychometric properties of the Center for Epidemiologic Studies Depression Scale (CES-D) in French Clinical and Non-Clinical Adults. Epidemiology and Public Health [Revue d'Épidémiologie et de Santé Publique] 59 (5):327-340

22. Kobor A, Takacs A, Urban R (2013) The Bifactor model of the Strenghts and Difficulties Questionnaire. European Journal of Psychological Assessment 29 (4):299-307

23. Meredith W (1993) Measurement invariance, factor analysis and factorial invariance. Psychometrika 58 (4):525-543

24. Millsap RE (2011) Statistical approaches to measurement invariance. Taylor & Francis, New York,NY

25. d'Acremont M, Van der Linden M (2008) Confirmatory factor analysis of the Strengths and Difficulties Questionnaire in a community of French-speaking adolescents. European Journal of Psychological Assessment 24 (1):1-8

26. Caci H, Morin AJS, Tran A (2014) Prevalence and correlates of Attention Deficit/Hyperactivity Disorder in

adults from a French community sample. Journal of Nervous and Mental Disease 202 (4):324-332

27. Muthén LK, Muthén BO (2012) Mplus User's Guide. Seventh Edition. Muthén & Muthén, Los Angeles, CA

28. Beauducel A, Herzberg PY (2006) On the Performance of Maximum Likelihood Versus Means and Variance Adjusted Weighted Least Squares Estimation in CFA. Structural Equation Modeling 13 (2):186-203

29. Finney SJ, DiStefano C (2006) Non-normal and categorical data in structural equation modeling. In: Hancock GR, Mueller RO (eds) Structural Equation Modeling: A Second Course. IAP, Greenwich,CT, pp 269-314

30. Flora DB, Curran PJ (2004) An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. Psychological Methods 9 (4):466-491

31. Forero CG, Maydeu-Olivares A, Gallardo-Pujol D (2009) Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. Structural Equation Modeling 16 (4):625-641

32. Muthén BO, du Toit SHC, Spisic D (1997) Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Mplus Technical Reports (http://gseis.ucla.edu/faculty/muthen/articles/Article_075.pdf).

33. Millsap RE, Tein JY (2004) Assessing Factorial Invariance in Ordered-Categorical Measures. Multivariate Behavioral Research 39 (3):479-515

34. Hu LT, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling 6 (1):1-55

35. Yu CY (2002) Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes. . University of California, Los Angeles,CA

36. Asparouhov T, Muthén BO (2006) Robust chi-square difference testing with mean and variance adjusted test statistics. Los Angeles, CA: Muthén and Muthén; http://www.statmodel.com/examples/webnote.shtml - web10.

37. Muthén BO (2004) Mplus Technical Appendices. Los Angeles, CA: Muthén and Muthén; http://www.statmodel.com/techappen.shtml.

38. Bollen KA (1989) Structural Equations with Latent Variables. Wiley, New York

39. Morin AJS, Madore I, Morizot J, Boudrias J-S, Tremblay M (2009) The Workplace Affective Commitment Multidimensional Questionnaire: Factor structure and Measurement Invariance. International Journal of Psychology Research 4 (3-4):307-344

40. Rensvold RB, Cheung GW (1998) Testing measurement model for factorial invariance: A systematic

approach. Educational and Psychological Measurement 58 (6):1017-1034

41. Chen FF (2007) Sensitivity of goodness of fit indexes to lack of measurement invariance. Structural Equation Modeling 14 (3):464-504

42. Cheung GW, Rensvold RB (2002) Evaluating goodness-of fit indexes for testing measurement invariance. Structural Equation Modeling 9 (2):233-255

43. McDonald RP (1970) The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. Bristish Journal of Mathematical and Statistical Psychology 23 (1):1-21

44. Sijtsma K (2009) On the use, misuse, and the very limited usefulness of Cronbach's alpha [Introduction to a special issue]. Psychometrika 74 (1):107-120

45. Le Corff Y, Toupin J ((Ahead of print)) Overt Versus Covert Conduct Disorder Symptoms and the Prospective Prediction of Antisocial Personality Disorder. Journal of Personality Disorders. doi:doi: 10.1521/pedi_2013_27_074

46. Connor DF, Glatt SJ, Lopez ID, Jackson D, Melloni RHJ (2002) Psychopharmacology and aggression. I: A meta-analysis of stimulant effects on overt/covert aggression-related behaviors in ADHD. Journal of the American Academy of Child and Adolescent Psychiatry 41 (3):253-261

47. Loeber R, Schmaling KB (1985) Empirical evidence for overt and covert patterns of antisocial conduct problems: a metaanalysis. Journal of Abnormal Child Psychology 13 (2):337-353

48. Arseneault L, Kim-Cohen J, Taylor A, Caspi A, Moffit TE (2005) Psychometric evaluation of 5- and 7-year-old children's self-reports of conduct problems. Journal of Abnormal Child Psychology 33 (5):537-550

49. Simonoff E, Pickles A, Hewitt J, Silberg J, Rutter M, Loeber R, Meyer J, Neale M, Eaves L (1995) Multiple raters of disruptive child behavior: using a genetic strategy to examine shared views and bias. Behavioral Genetics 25 (4):311-326

50. Rubin KH, Copland RJ, Bowker JC (2009) Social withdrawal in childhood. Annual Review of Psychology 60:141-171

51. Wang JM, Rubina KH, Laursenb B, Booth-LaForcec C, Rose-Krasnord L (2013) Preference-for-Solitude and Adjustment Difficulties in Early and Late Adolescence. Journal of Clinical Child & Adolescent Psychology 42 (6):834-842

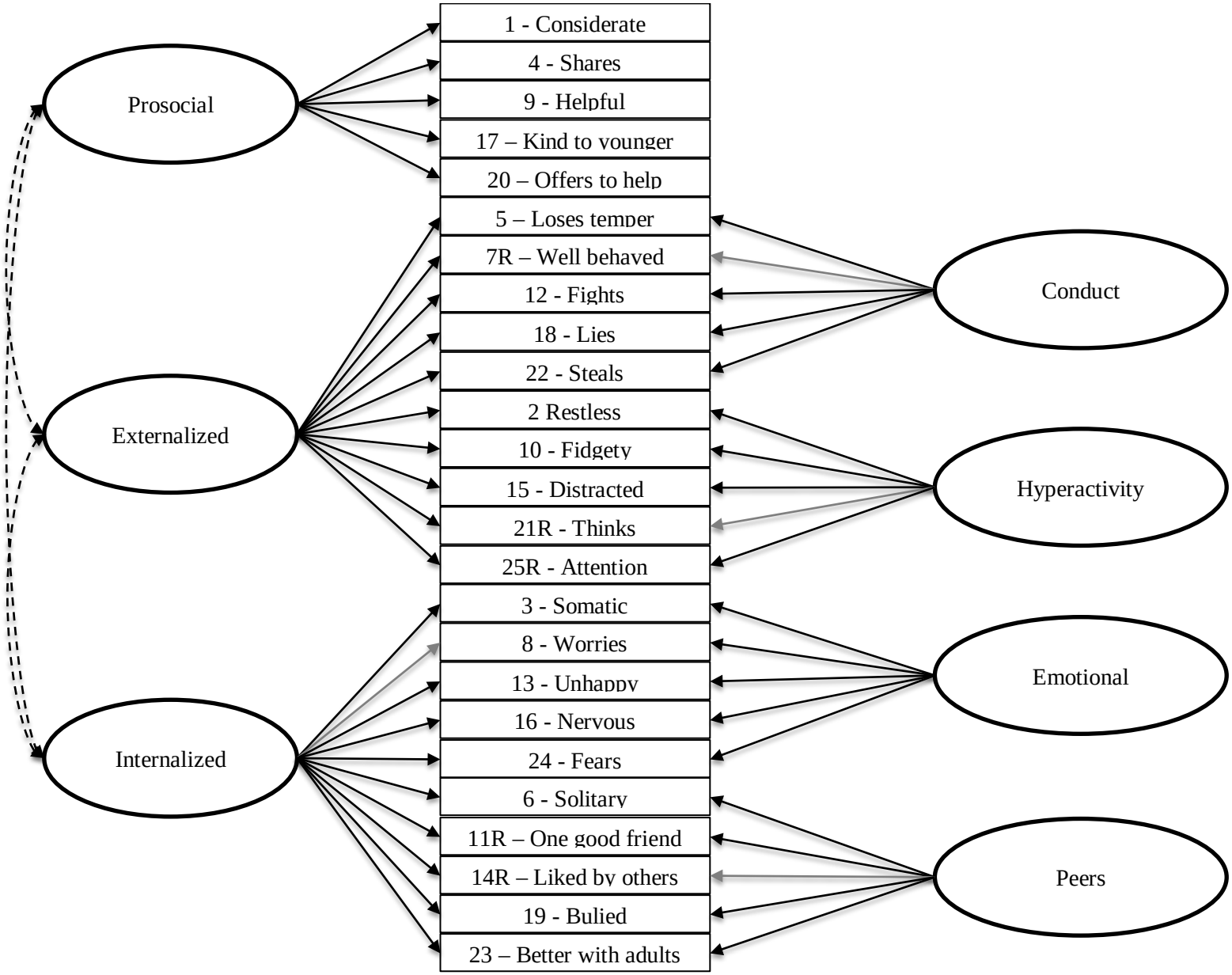

Figure I – Diagram for the final retained bifactor Model M9 : loadings are represented as solid lines with an arrow (non-significant loadings are in gray) and factor correlations are represented as broked lines with arrows at both ends (see numerical values in Table 3).

Table 1.

Fit Indices for the Alternative Measurement Models.

Model		χ2	df	CFI	TLI	RMSEA	RMSEA 90% CI	Δχ2	Δdf	ΔCFI	ΔTLI	ΔRMSEA
M1.	1-factor model (Global)	3247.933*	275	.766	.744	.110	.107-.114					
M2.	2-factor model (S & D)	2795.792*	274	.801	.782	.102	.098-.105					
M3.	3-factor model (S, E, I)	1981.337*	272	.865	.851	.084	.081-.088					
M4.	Alternative 3-factor model	1649.682*	272	.891	.880	.075	.072-.079					
M5.	5-factor model	1349.631*	265	.914	.903	.068	.064-.071					
M6.	Higher-order (S & D)	1591.795*	270	.896	.884	.074	.071-.078					
M7.	Higher-order (S, E, I)	1387.147*	268	.912	.901	.069	.065-.072					
M8.	Bifactor (S & D)	1310.780*	254	.917	.902	.068	.065-.072					
M9.	Bifactor (S, E, I)	1080.672*	252	.935	.922	.061	.057-.065					
Measurement Invariance (M9) by gender groups												
MG1.	Configural invariance	1268.703*	504	.932	.919	.058	.054-.062	--	--	--	--	--
MG2.	Metric/Weak invariance	1305.754*	542	.937	.930	.056	.052-.060	64.290*	38	.005	.011	-.002
MG3.	Scalar/Strong invariance	1239.423*	560	.940	.935	.052	.048-.056	58.570*	56	.003	.005	-.004
MG4.	Strict invariance	1225.224*	585	.943	.942	.050	.046-.054	26.505	25	.003	.007	-.002
MG5.	Latent variances-covariances invariance	1199.467*	592	.946	.945	.048	.044-.052	11.852	7	.003	.003	-.002
MG6.	Latent means invariance	1392.166*	599	.929	.929	.055	.051-.058	110.909*	7	-.017	-.016	.007
Measurement Invariance (M9) by school level												
MS1.	Configural invariance	1579.005*	756	.927	.913	.060	.056-.065	—	—	—	—	—
MS2.	Metric/Weak invariance	1596.525*	832	.932	.926	.055	.051-.060	118.118*	76	.005	.013	-.005
MS3.	Scalar/Strong invariance	1611.085*	868	.934	.931	.054	.049-.058	45.037	36	.002	.005	-.001
MS4.	Strict invariance	1638.032*	904	.935	.935	.052	.048-.056	76.210*	36	.001	.004	-.002
MS5.	Latent variances-covariances invariance	1617.358*	918	.938	.939	.051	.046-.055	23.650	14	.003	.004	-.001
MS6.	Latent means invariance	1628.433*	932	.938	.940	.050	.046-.054	28.308*	14	.000	.001	-.001

Notes. S: Strengths; D=Difficulties; E=Externalizing disorders; I=Internalizing disorders; *p<0.01; χ^2: chi-square test of model fit and its associated degrees of freedom (*df*); CFI: Comparative Fit Index; TLI: Tucker-Lewis Index; RMSEA: Root Mean Square Error of Approximation and its 90% Confidence Interval (CI); Δ change relative to the previous model in the sequence; MDΔχ^2: chi-square difference test calculated with the Mplus DIFFTEST function for the robust weighted least square estimator (WLSMV). The fact that WLSMV χ^2 values are not exact, but "*estimated*" as the closest integer necessary to obtain a correct *p*-value explains the fact that the χ^2 and resulting CFI values can be non-monotonic with model complexity.

Table 2.

Original First-Order 5-factor model (M5)

Item	Item content	Strengths/ Prosocial	Conduct Problems	Hyperactivity-Inattention	Emotional Symptoms	Peer Problems	Residual ($1-R^2$)
1	Considerate	0.798					0.363
4	Shares	0.682					0.535
9	Helpful	0.804					0.353
17	Kind to younger	0.764					0.417
20	Offers to help	0.827					0.316
5	Loses temper		0.777				0.396
7-R	Well behaved		0.630				0.604
12	Fights		0.820				0.327
18	Lies		0.740				0.453
22	Steals		0.540				0.709
2	Restless			0.951			0.096
10	Fidgety			0.908			0.175
15	Distracted			0.742			0.449
21-R	Thinks			0.791			0.374
25-R	Attention			0.709			0.497
3	Somatic				0.585		0.658
8	Worries				0.608		0.630
13	Unhappy				0.905		0.181
16	Nervous				0.770		0.406
24	Fears				0.779		0.393
6	Solitary					0.413	0.829
11-R	One good friend					0.779	0.394
14-R	Liked by others					0.928	0.139
19	Bullied					0.643	0.587
23	Better with adults					0.248	0.938
Scale-Score reliability:	ω	0.883	0.832	0.914	0.854	0.758	
Factor Correlations		Factor 1	Factor 2	Factor 3	Factor 4		
	Factor 2	-0.663					
	Factor 3	-0.511	0.748				
	Factor 4	-0.162	0.234	0.290			
	Factor 5	-0.734	0.507	0.370	0.513		

Note: All coefficients are significant ($p \leq 0.05$).

Table 3.

Final Retained Bifactor Model (M9).

Item	Item content	First-Order Strengths/ Prosocial	S-Factor Conduct Problems	S-Factor Hyperactivity-Inattention	S-Factor Emotional Symptoms	S-Factor Peer Problems	G-Factor Internalizing Disorders	G-Factor Externalizing Disorders	Residual ($1-R^2$)
1	Considerate	0.800							0.359
4	Shares	0.679							0.539
9	Helpful	0.802							0.357
17	Kind to younger	0.762							0.419
20	Offers to help	0.830							0.312
5	Loses temper		0.295					0.719	0.395
7-R	Well behaved		*-0.119*					0.613	0.610
12	Fights		0.293					0.761	0.334
18	Lies		0.492					0.664	0.318
22	Steals		0.439					0.467	0.589
2	Restless			0.622				0.689	0.138
10	Fidgety			0.758				0.649	0.005
15	Distracted			0.188				0.706	0.466
21-R	Thinks			*-0.118*				0.791	0.361
25-R	Attention			-0.161				0.717	0.460
3	Somatic				0.468		0.260		0.714
8	Worries				0.780		*0.082*		0.386
13	Unhappy				0.551		0.593		0.345
16	Nervous				0.718		0.299		0.395
24	Fears				0.767		0.280		0.333
6	Solitary					0.570	0.305		0.582
11-R	One good friend					0.313	0.740		0.355
14-R	Liked by others					*0.123*	0.919		0.141
19	Bullied					0.277	0.596		0.568
23	Better with adults					0.707	0.127		0.484
Scale-Score reliability: ω		0.883	0.544	0.705	0.832	0.650	0.804	0.926	
Factor Correlations		Factor 1	G-Factor 1						
	G-Factor 1	-0.730**							
	G-Factor 2	-0.635**	0.561**						

Non-significant coefficients are marked in italics. All other coefficients are significant ($p \leq 0.05$).

Online Supplements for:

Investigation of a bifactor model of the Strengths and Difficulties Questionnaire

Hervé CACI, Alexandre J. S. MORIN, & Antoine TRAN

European Child and Adolescent Psychiatry

Corresponding author:

Hervé CACI, MD, PhD
Hôpitaux Pédiatrique de Nice – CHU Lenval
57, avenue de la Californie
06200 Nice
Tel : +33 (0)4 92 03 08 71
Fax : +33 (0)4 92 03 05 58
Email : caci.h@pediatrie-chulenval-nice.fr

Table S1

Studies that investigated the factor structure of the SDQ (self-report version).

Study	N (boys/girls)	Age/grade range	Language (Country)	Method	Model	Fit
Altendorder-Kling et al. (2007)	2,529 (1,237/1,292)	11-18	German (Austria)	PCA (Varimax)	5-component (confirmatory)	—
Capron et al. (2007)	1,400 (692/708)	12-13	French (France)	PCA (Varimax)	5-component (confirmatory)	—
Di Riso et al. (2010)	1,394 (712/682)	8-10	Italian (Italy)	CFA	5-factor	Good
				CFA	3-factor (8-item prosocial)	Acceptable
Du et al. (2008)	690 (326/364)	11-17	Chinese (China)	PCA (Varimax)	5-component (confirmatory)	—
Essau et al. (2012)	2,418 (1,001/1,417)	12-17	UK, Germany, Cyprus, Sweden & Italy	CFA	5-factor	Good
				CFA	3-factor (8-item prosocial)	Acceptable
Giannakopoulos et al. (2009)	1,194 (479/715)	11-17	Greek (Greece)	CFA	5-factor	Poor
Goodman et al. (2010)	7,678 (unspecified)	11-16	English (UK)	CFA	5-factor	Poor
				CFA	5-factor, 2 HO (Int, Ext)	Poor
				CFA	3-factor	Poor
Goodman et al. (2001)	3,983 (unspecified)	11-15	English (UK)	PCA (Varimax)	5-component (confirmatory)	—
Koskelainen et al. (2001)	1,458 (733/725)	13-17	Finnish (Finland)	PCA (Varimax)	5-component (confirmatory)	—
					3-component (exploratory)	—
Liu et al. (2013)	2672 (1400/1272)	8 and above	Chinese (China)	EFA (Varimax)	5-factor (exploratory)	—
Lundh et al. (2008)	926 (unspecified)	14-15	Swedish (Sweden)	PCA (Varimax)	5-component (confirmatory)	—
Mansbach-Kleinfeld (2010)	611 (317/294)	14-16	Hebrew (Israel)	PCA (Varimax)	5-component (exploratory)	—
				CFA	5-factor	no CFI
Mellor et al. (2007)	914 (425/489)	7-17	English (Australia)	CFA	5-factor, 1 HO (difficulties)	Poor
Muris et al. (2003)	562 (254/308)	9-15	Dutch (Netherlands)	PCA (Oblimin)	5-component (exploratory)	—
Muris et al. (2004)	1111 (551-560)	8-13	Dutch (Netherlands)	PCA (Oblimin)	4 and 5 factors (exploratory)	—
	490 (unspecified)	8-10	Dutch (Netherlands)	PCA (Oblimin)	4 and 5 factors (exploratory)	—
	621 (unspecified)	11-13	Dutch (Netherlands)	PCA (Oblimin)	4 and 5 factors (exploratory)	—
Percy et al. (2007)	1st half split of 3,753 (1,802/1,951)	12	English (Ireland)	EFA (Promax)	3 and 5 factors (exploratory)	—
				CFA	5-factor	Poor
				CFA	3-factor	Poor
	2nd half split of 3,753 (1,802/1,951)	12	English (Ireland)	EFA (Promax)	3 and 5 factors (exploratory)	—
				CFA	5-factor	Poor
Perera et al. (2013)	1,180 (535/645)	12-16	Sinhalese (Sri Lanka)	PCA (Varimax)	5-component (confirmatory)	—
Poulou (2013)	559 (294/265)	12-14	Greek (Greece)	CFA	5-factor	Acceptable
Richter et al. (2011)	5,379 ethnic Norwegian (2,663/2,712)	15-16	Norwegian (Norway)	CFA	5-factor	Good
	516 Pakistani	15-16	Norwegian (Norway)	CFA	5-factor	Good
	349 other ethnic minorities	15-16	Norwegian (Norway)	CFA	5-factor	Acceptable
Ronning et al. (2004)	4,167 (2,150/2,017)	11-16	Norwegian (Norway)	CFA	5-factor	Poor
Ruchkin et al. (2007)	2,892 (1,226/1,666)	12-17	Russian (Russia)	CFA	5-factor	Not reported
Ruchkin et al. (2008)	4,671 urban sample	6th- 10th grade	English (USA)	CFA	5-factor	Acceptable
					Post-hoc 3-factor	Good
	First-half (urban)	6th- 10th grade	English (USA)	CFA	5-factor	Acceptable
					Post-hoc 3-factor	Good
	Second-half (urban)	6th- 10th grade	English (USA)	CFA	5-factor	Acceptable
					Post-hoc 3-factor	Good
	937 (450/487) suburban sample	6th- 10th grade	English (USA)	CFA	5-factor	Good
					Post-hoc 3-factor	Acceptable
	First-half (suburban)	6th- 10th grade	English (USA)	CFA	5-factor	Good
					Post-hoc 3-factor	Acceptable
	Second-half (suburban)	6th- 10th grade	English (USA)	CFA	5-factor	Good
					Post-hoc 3-factor	Acceptable
Van de Looij-Jansen et al. (2011)	7,921 (unspecified)		Dutch (Netherlands)	CFA	5-factor	Poor

Study	N (boys/girls)	Age/grade range	Language (Country)	Method	Model	Fit
van Roy et al. (2008)	8,320	10-13	Norwegian (Norway)	CFA	5-factor	Good
	8,582	13-16	Norwegian (Norway)	CFA	5-factor	Good
	9,367	16-19	Norwegian (Norway)	CFA	5-factor	Good
Yao et al. (2009)	1,132	11-18	Chinese (China)	CFA	5-factor	Acceptable
					5-factor, 1 HO (difficulties)	Acceptable
	Younger (unspecified)	11-14	Chinese (China)	CFA	5-factor	Good
					5-factor, 1 HO (difficulties)	Acceptable
	Older (unspecified)	15-18	Chinese (China)	CFA	5-factor	Poor
					5-factor, 1 HO (difficulties)	Poor
	561 boys	—	Chinese (China)	CFA	5-factor	Acceptable
					5-factor, 1 HO (difficulties)	Acceptable
	574 girls	—	Chinese (China)	CFA	5-factor	Acceptable
					5-factor, 1 HO (difficulties)	Acceptable

Note: PCA: Principal Component Analysis; EFA: Exploratory Factor Analysis; CFA: Confirmatory Factor Analysis; HO: Higher-order factor; Int: Internalizing disorders; Ext: Externalizing disorders; Good fit: CFI>.95 and RMSEA<.06; Acceptable fit: CFI>.90 and RMSEA<.08; Poor fit: CFI<.90 or RMSEA>.08. EFA and PCA were labeled either exploratory when solutions with different number of factors were contrasted on the basis of various criteria (e.g., parallel analysis, scree test) and confirmatory when only the expected number of factor were estimated.

Table S2

Studies that investigated the factor structure of the SDQ (parent version).

Study	N (boys/girls)	Age/grade range	Language (Country)	Method	Model	Fit
Becker et al. (2004)	543 (396/147)	5-17	German (Germany)	CFA	5-factor	Poor
	1,686 (mix clinical & community)	5-17	German (Germany)	PCA (Varimax)	5-component (confirmatory)	—
Becker et al. (2006)	1,459 (1,222/231)	6-18	10 European countries	PCA (Varimax)	5-component (confirmatory)	—
Björnsdotter et al. (2013)	457 online	10-13	Swedish (Sweden)	CFA	5-factor	Good
Dickey & Blumberg (2004)	9,577	4-17	English (USA)	PCA (Varimax)	5-component (confirmatory)	—
	4,773 (1st half sample)	4-17	English (USA)	EFA (Promax) 24 items	3-factor (confirmatory)	—
	4,804 (2nd half sample)	4-17	English (USA)	CFA 24 items	3-factor (8-item prosocial)	Good
Du et al. (2008)	1,965 (950/1,015)	3-17	Chinese (China)	PCA (Varimax)	5-component (confirmatory)	—
Gharehbaghy et al. (2009)	413 (193/220)	10-12	Persian (Iran)	PCA	3- and 5-component (exploratory)	—
Gomez-Beneyto et al. (2013)	3,253 (1st random half)	4-15	Spanish (Spain)	EFA (Promin)	3- and 5-factor (exploratory)	—
	3,253 (2nd random half)	4-15	Spanish (Spain)	CFA	5-factor	Good
Goodman et al. (2010)	18,222	5-16	English (UK)	CFA	5-factor	Poor
				CFA	5-factor, 2 HO (Int, Ext)	Acceptable
				CFA	3-factor	Poor
Goodman et al. (2001)	9,998	5-15	English (UK)	PCA (Varimax)	5-component (confirmatory)	—
Hawes et al. (2004)	706 boys	4-9	English (Australia)	PCA (Oblimin)	5-component (confirmatory)	—
	653 girls	4-9	English (Australia)	PCA (Oblimin)	5-component (confirmatory)	—
He et al. (2013)	3,333 girls	13-18	English (USA)	CFA	5-factor	Good
	3,150 boys	13-18	English (USA)	CFA	5-factor	Acceptable
	758 Hispanic	13-18	English (USA)	CFA	5-factor	Poor
	1,097 Black	13-18	English (USA)	CFA	5-factor	Acceptable
	2,457 White	13-18	English (USA)	CFA	5-factor	Good
	5,139	13-16	English (USA)	CFA	5-factor	Acceptable
	1,344	17-18	English (USA)	CFA	5-factor	Acceptable
	2,143 low SES	13-18	English (USA)	CFA	5-factor	Acceptable
	4,340 high SES	13-18	English (USA)	CFA	5-factor	Good
Hill et al. (2007)	505 (269/236)	6.11 ± .65	English (USA)	CFA	5-factor	Poor
Liu et al. (2013)	3546 (1832/1714)	8 and above	Chinese (China)	EFA (Varimax)	4-factor (exploratory)	—
Mansbach-Kleinfeld (2010)	553 (298/294)	14-16	Hebrew (Israel)	PCA (Varimax)	5-component (exploratory)	—
				CFA	5-factor	no CFI
Matsuishi et al. (2008)	2,899 (1,463/1,436)	4-12	Japanese (Japan)	PCA (Varimax)	5-component (confirmatory)	—
McCrory & Layte (2012)	8,514	9	English (Ireland)	CFA	5-factor	Poor
					5-factor, 1 HO (difficulties)	Poor
					3-factor	Poor
Mellor et al. (2007)	914 (425/489)	7-17	English (Australia)	CFA	5-factor, 1 HO (difficulties)	Poor
Mieloo et al. (2014)	4,325 Dutch	4-6 (2nd grade)	Dutch (Netherlands)	PCA (Oblimin)	5-component (confirmatory)	—
	450 Surinamese	4-6 (2nd grade)	Dutch (Netherlands)	PCA (Oblimin)	5-component (confirmatory)	—
	190 Antillean or Aruban	4-6 (2nd grade)	Dutch (Netherlands)	PCA (Oblimin)	5-component (confirmatory)	—
	526 Turkish	4-6 (2nd grade)	Dutch (Netherlands)	PCA (Oblimin)	5-component (confirmatory)	—
	467 Moroccan	4-6 (2nd grade)	Dutch (Netherlands)	PCA (Oblimin)	5-component (confirmatory)	—
Moriwaki & Kamio (2014)	24,519 (12,472/12,047)	7-15	Japanese (Japan)	EFA	5-factor (confirmatory)	—
				CFA	5-factor	Poor
Muris et al. (2003)	562 (254/308)	9-15	Dutch (Netherlands)	PCA (Oblimin)	5-component (confirmatory)	—
Niclasen et al. (2012)	27,611 younger girls	5-7	Danish (Denmark)	PCA (Promax)	5-component (confirmatory)	—
	28,920 younger boys	5-7	Danish (Denmark)	PCA (Promax)	5-component (confirmatory)	—
	3,237 older girls	10-12	Danish (Denmark)	PCA (Promax)	5-component (confirmatory)	—
	3,322 older boys	10-12	Danish (Denmark)	PCA (Promax)	5-component (confirmatory)	—

Study	N (boys/girls)	Age/grade range	Language (Country)	Method	Model	Fit
Niclasen et al. (2013)	27,611 younger girls	5-7	Danish (Denmark)	CFA	5-factor	Poor
				CFA	5-factor, 1 HO (difficulties)	Poor
				CFA	5-factor, 2 HO (Int, Ext)	Poor
	28,920 younger boys	5-7	Danish (Denmark)	CFA	5-factor	Poor
				CFA	5-factor, 1 HO (difficulties)	Poor
				CFA	5-factor, 2 HO (Int, Ext)	Poor
	3,237 older girls	10-12	Danish (Denmark)	CFA	5-factor	Acceptable
				CFA	5-factor, 1 HO (difficulties)	Acceptable
				CFA	5-factor, 2 HO (Int, Ext)	Acceptable
	3,322 older boys	10-12	Danish (Denmark)	CFA	5-factor	Acceptable
				CFA	5-factor, 1 HO (difficulties)	Acceptable
				CFA	5-factor, 2 HO (Int, Ext)	Acceptable
Palmieri & Smith (2007)	733 (342/391)	4-10	English (USA)	CFA	5-factor	Good
				CFA	5-factor, 1 HO (difficulties)	Good
Rodriguez-Hernandez et al. (2012)	530	7-10	Spanish (Canary Islands)	EFA	5-factor (exploratory)	—
Rothenberger et al. (2008)	2,406	7-17	German (Germany)	PCA (Varimax)	5-component (confirmatory)	—
				CFA	5-factor	Not reported
Sanne (2009)	3,264 (1st random split)	7-9 (Grades 2-4)	Norwegian (Norway)	EFA (Promax)	5 factor (exploratory)	—
	3,166 (2nd random split)	7-9 (Grades 2-4)	Norwegian (Norway)	CFA	5-factor	Acceptable
				CFA	3-factor (8-item prosocial)	Poor
Smedje et al. (1999)	900 (460/440)	6-8 & 10	Swedish (Sweden)	PCA (Varimax)	5-component (confirmatory)	—
van Leeuwen et al. (2006)	532 (250/282)	4-5 (Preschool, 67%) & 6-7 (Primary school, 33%)	Dutch (Flanders)	EFA (Oblimin)	3- and 5-factor (exploratory)	—
				CFA	5-factor	Poor
				CFA	3-factor (8-item prosocial)	Poor
	1,086 (532/554)	4-5 (Preschool, 79%) & 6-7 (Primary school, 21%)	Dutch (Flanders)	EFA (Oblimin)	3- and 5-factor (exploratory)	—
				CFA	5-factor	Poor
				CFA	3-factor (8-item prosocial)	Poor
van Roy et al. (2008)	6,645 (unspecified)	10-13	Norwegian (Norway)	CFA	5-factor	Good
Woerner et al. (2004)	930 (467/463)	6-16	German (Germany)	PCA(Varimax)	5-component (confirmatory)	—

Note: PCA: Principal Component Analysis; EFA: Exploratory Factor Analysis; CFA: Confirmatory Factor Analysis; HO: Higher-order factor; Int: Internalizing disorders; Ext: Externalizing disorders; Good fit: CFI>.95 and RMSEA<.06; Acceptable fit: CFI>.90 and RMSEA<.08; Poor fit: CFI<.90 or RMSEA>.08. EFA and PCA were labeled either exploratory when solutions with different number of factors were contrasted on the basis of various criteria (e.g., parallel analysis, scree test) and confirmatory when only the expected number of factor were estimated.

Table S3

Studies that investigated the factor structure of the SDQ (teacher version).

Study	N (boys/girls)	Age/grade range	Language (Country)	Method	Model	Fit
Becker et al. (2004)	543 (396/147)	5-17	German (Germany)	PCA (Varimax)	5-component (confirmatory)	—
Capron et al. (2007)	1,400 (692/708)	12-13	French (France)	PCA (Varimax)	5-component (confirmatory)	—
d'Acremont et al. (2008)	557 (278/279)	13-18	French (Switzerland)	CFA	5-factor	Poor
Du et al. (2008)	1,965 (950/1,015)	3-17	Chinese (China)	PCA (Varimax)	5-component (confirmatory)	—
Gharehbaghy et al. (2009)	413 (193/220)	10-12	Persian (Iran)	PCA	3- and 5-component (exploratory)	—
Goodman et al. (2010)	14,263	5-16	English (UK)	CFA	5-factor	Poor
				CFA	5-factor, 2 HO (Int, Ext)	Acceptable
				CFA	3-factor	Poor
Goodman et al. (2001)	7,313	5-15	English (UK)	PCA (Varimax)	5-component (confirmatory)	—
Hayes (2007)	905 boys	5-10	English (Australia)	PCA (Varimax)	5-component (confirmatory)	—
	1023 girls	5-10	English (Australia)	PCA (Varimax)	5-component (confirmatory)	—
Hill et al. (2007)	676 (350/326)	6.12 ± .65	English (USA)	CFA	5-factor	Poor
Kashala et al. (2005)	1,187 (502/685)	7-9	French (Congo)	PCA (Varimax)	5-component (confirmatory)	—
Liu et al. (2013)	3669 (1918/1751)	8 and above	Chinese (China)	EFA (Varimax)	4-factor (exploratory)	—
Mellor et al. (2007)	914 (425/489)	7-17	English (Australia)	CFA	5-factor, 1 HO (difficulties)	Poor
Mieloo et al. (2014)	4,314 Dutch	4-6 (2nd grade)	Dutch (Netherlands)	PCA (Oblimin)	5-component (confirmatory)	—
	591 Surinamese	4-6 (2nd grade)	Dutch (Netherlands)	PCA (Oblimin)	5-component (confirmatory)	—
	318 Antillean or Aruban	4-6 (2nd grade)	Dutch (Netherlands)	PCA (Oblimin)	5-component (confirmatory)	—
	739 Turkish	4-6 (2nd grade)	Dutch (Netherlands)	PCA (Oblimin)	5-component (confirmatory)	—
	776 Moroccan	4-6 (2nd grade)	Dutch (Netherlands)	PCA (Oblimin)	5-component (confirmatory)	—
Moriwaki & Kamio (2014)	7,977 (4,010/3,967)	7-15	Japanese (Japan)	EFA	5-factor (confirmatory)	—
				CFA	5-factor	Poor
Niclasen et al. (2012)	1,291 younger girls	5-7	Danish (Denmark)	PCA (Promax)	5-component (confirmatory)	—
	1,272 younger boys	5-7	Danish (Denmark)	PCA (Promax)	5-component (confirmatory)	—
	2,805 older girls	10-12	Danish (Denmark)	PCA (Promax)	5-component (confirmatory)	—
	2,790 older boys	10-12	Danish (Denmark)	PCA (Promax)	5-component (confirmatory)	—
Niclasen et al. (2013)	1,291 younger girls	5-7	Danish (Denmark)	CFA	5-factor	Acceptable
				CFA	5-factor, 1 HO (difficulties)	Acceptable
				CFA	5-factor, 2 HO (Int, Ext)	Acceptable
	1,272 younger boys	5-7	Danish (Denmark)	CFA	5-factor	Acceptable
				CFA	5-factor, 1 HO (difficulties)	Acceptable
				CFA	5-factor, 2 HO (Int, Ext)	Acceptable
	2,805 older girls	10-12	Danish (Denmark)	CFA	5-factor	Good
				CFA	5-factor, 1 HO (difficulties)	Good
				CFA	5-factor, 2 HO (Int, Ext)	Good
	2,790 older boys	10-12	Danish (Denmark)	CFA	5-factor	Good
				CFA	5-factor, 1 HO (difficulties)	Good
				CFA	5-factor, 2 HO (Int, Ext)	Good
Rodriguez-Hernandez et al. (2012)	595 (309/286)	7-10	Spanish (Canary Islands)	EFA	5-factor (exploratory)	—
				CFA	5-factor	Acceptable
Ruckin et al. (2012)	538 (259/277)	12-17	Russian (Russia)	CFA	5-factor	Poor
Sanne (2009)	4,516 (1st random split)	7-9	Norwegian (Norway)	EFA (Promax)	5-factor (exploratory)	—
	4,483 (2nd random split)	7-9	Norwegian (Norway)	CFA	5-factor	Acceptable
				CFA	3-factor (8-item prosocial)	Poor
Tobia et al. (2013)	1000 (random sample 1)	3-15	Italian (Italy)	EFA (Geomin)	5-factor (confirmatory)	—
	2302 (random sample 2)	3-15	Italian (Italy)	CFA	5-factor	Acceptable
	2302 (random sample 2)	3-15	Italian (Italy)	CFA	5-factor, 2 HO (Int, Ext)	Acceptable
	2302 (random sample 2)	3-15	Italian (Italy)	CFA	post-hoc 5-factor	Good
van Leeuwen et al. (2006)	512 (240/272)	4-5 (Preschool, 67%) & 6-7 (Primary school, 33%)	Dutch (Flanders)	EFA (Oblimin)	3 and 5 factors (confirmatory)	—
				CFA	5-factor	Acceptable

Study	N (boys/girls)	Age/grade range	Language (Country)	Method	Model	Fit
				CFA	3-factor (8-item prosocial)	Poor
	1,049 (514/465)	4-5 (Preschool, 79%) & 6-7 (Primary school, 21%)	Dutch (Flanders)	CFA	5-factor	Poor
Zwirs	684 Dutch (352/332)	6-10	Dutch (Netherlands)	CFA	5-factors	Not reported
	702 Moroccan (378/324)			CFA	5-factors	Not reported
	434 Turkish (212/222)			CFA	5-factors	Not reported
	365 Surinamese (168/197)			CFA	5-factors	Not reported

Note: PCA: Principal Component Analysis; EFA: Exploratory Factor Analysis; CFA: Confirmatory Factor Analysis; HO: Higher-order factor; Int: Internalizing disorders; Ext: Externalizing disorders; Good fit: CFI>.95 and RMSEA<.06; Acceptable fit: CFI>.90 and RMSEA<.08; Poor fit: CFI<.90 or RMSEA>.08. EFA and PCA were labeled either exploratory when solutions with different number of factors were contrasted on the basis of various criteria (e.g., parallel analysis, scree test) and confirmatory when only the expected number of factor were estimated.

References Used in the Online Supplements.

1. Altendorder-Kling U, Ardelt-Gattinger E, Thun-Hohenstein L (2007) [The self-report version of the Strengths and Difficulties Questionnaire in an Austrian field sample]. Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie 35 (4):265-271

2. Becker A, Steinhausen HC, Baldursson G, Dalsgaard S, Lorenzo MJ, Ralston SJ, Döpfner M, Rothenberger A, Group AS (2006) Psychopathological screening of children with ADHD: Strengths and Difficulties Questionnaire in a pan-European study. European Child and Adolescent Psychiatry 15 (Suppl.1):i56-i62

3. Becker A, Woerner W, Hasselhorn M, Banaschewski T, Rothenberger A (2004) Validation of the parent and teacher SDQ in a clinical sample. European Child and Adolescent Psychiatry 13 (Suppl.2):11-16

4. Björnsdotter A, Enebrink P, Ghaderi A (2013) Psychometric properties of online administered parental Strengths and Difficulties Questionnaire (SDQ) and normative data based on combined online and paper-and-pencil administration. Child and Adolescent Psychiatry and Mental Health 7 (1):40

5. Capron C, Thérond C, Duyme M (2007) Psychometric properties of the French version of the Self-report and Teacher Strengths and Difficulties Questionnaire (SDQ). European Journal of Psychological Assessment 23 (2):79-88

6. d'Acremont M, Van der Linden M (2008) Confirmatory factor analysis of the Strengths and Difficulties Questionnaire in a community of French-speaking adolescents. European Journal of Psychological Assessment 24 (1):1-8

7. Di Riso 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

8. Dickey WC, Blumberg SJ (2004) Revisiting the factor structure of the Strengths and Difficulties Questionnaire: United States, 2001. Journal of the American Academy of Child and Adolescent Psychiatry 43 (9):1159-1167

9. Du Y, Kou J, Coghill D (2008) The validity, reliability and normative scores of the parent, teacher and self report versions of the Strengths and Difficulties Questionnaire in China. Child and Adolescent Psychiatry and Mental Health. doi:10.1186/1753-2000-2-8

10. Essau CA, Olaya B, Anastassiou-Hadjicharalambous X, Pauli G, Gilvarry C, Bray D, O'callaghan J, Ollendick TH (2012) Psychometric properties of the Strengths and Difficulties Questionnaire from five European countries. International Journal of Methods in Psychiatric Research 21 (3):232-245

11. Gharehbaghy F, Aguilar-Vafaie M (2009) Psychometric properties of Persian parent and teacher versions of the Strenghts and Difficulties Questionnaire in a sample of Iranian children. Iranian Journal of Psychiatry and Clinical Psychology 15 (3):231-241

12. 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 26 (8):20

13. Gómez-Beneyto M, Nolasco A, Moncho J, Pereyra-Zamora P, Tamayo-Fonseca N, Munarriz M, Salazar J, Tabarés-Seisdedos R, Girón M (2013) Psychometric behaviour of the strengths and difficulties questionnaire (SDQ) in the Spanish national health survey 2006. BMC Psychiatry 13:95. doi:10.1186/1471-244X-13-95

14. Goodman A, Lamping DL, Ploubidis GB (2010) When to use broader internalising and externalising subscales instead of the hypothesized five subscales on the Strengths and Difficulties Questionnaire (SDQ): data from British parents, teachers and children. Journal of Abnormal Child Psychology 38 (8):1179-1191

15. Goodman R (2001) Psychometric properties of the strengths and difficulties questionnaire. Journal of the American Academy of Child and Adolescent Psychiatry 40 (11):1337-1345

16. Hawes DJ, Dadds MR (2004) Australian data and psychometric properties of the Strengths and Difficulties Questionnaire. Australian and New Zealand Journal of Psychiatry 38 (8):644-651

17. Hayes L (2007) Problem behaviours in early primary school children: Australian normative data using the Strengths and Difficulties Questionnaire. Australian and New Zealand Journal of Psychiatry 41 (3):231-238

18. He JP, Burstein M, Schmitz A, Merikangas KR (2013) The Strengths and Difficulties Questionnaire (SDQ): the factor structure and scale validation in U.S. adolescents. Journal of Abnormal Child Psychology 41 (4):583-595

19. Hill CR, Hughes JN (2007) An examination of the convergent and discriminant validity of the Strengths and Difficulties Questionnaire. School Psychology Quaterly 22 (3):380-406

20. Hu LT, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling 6 (1):1-55

21. Kashala E, Elgen I, Sommerfelt K, Tylleskar T (2005) Teacher ratings of mental health among school children in Kinshasa, Democratic Republic of Congo. European Child and Adolescent Psychiatry 14 (4):208-215

22. 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 (3):180-185

23. Liu S-K, Chien Y-L, Shang C-Y, Lin C-H, Liu Y-C, Gau SS-F (2013) Psychometric properties of the Chinese version of Strengths and Difficulties Questionnaire. Comprehensive Psychiatry 54 (6):720-730

24. Lundh LG, Wangby-Lundh M, Bjärehed J (2008) Self-reported emotional and behavioral problems in Swedish 14 to 15-year-old adolescents: a study with the self-report version of the Strengths and Difficulties Questionnaire. Scandinavian Journal of Psychology 49 (6):523-533

25. Mansbach-Kleinfeld I, Apter A, Farbstein I, Levine SZ, Ponizovsky AM (2010) A population-based psychometric validation study of the strengths and difficulties questionnaire - hebrew version. Frontiers in Psychiatry 31 (1):151. doi:10.3389/fpsyt.2010.00151

26. Matsuishi T, Nagano M, Araki Y, Tanaka Y, Iwasaki M, Yamashita Y, Nagamitsu S, Izuka C, Ohya T, Shibuya K, Hara M, Matsuda K, Tsuda A, Kakuma T (2008) Scale properties of the Japanese version of the Strengths and Difficulties Questionnaire (SDQ): a study of infant and school children in community samples. Brain and Development 30 (6):410-415

27. McCrory C, Layte R (2012) Testing competing models of the Strenghts and Difficulties Questionnaire's (SDQ) factor structure for the parent-informant instrument. Personality and Individual Differences 52 (8):882-887

28. Mellor D, Stokes M (2007) The factor structure of the Strengths and Difficulties Questionnaire. European Journal of Psychological Assessment 23 (2):105-112

29. Mieloo CL, Bevaart F, Donker MC, van Oort FV, Raat H, Jansen W (2014) Validation of the SDQ in a multi-ethnic population of young children. European Journal of Public Health 24 (1):26-32

30. Muris P, Meesters C, Eiijkelenboom A, Vincken M (2004) The self report version of the Strengths and Difficulties Questionnaire: its psychometric properties in 8-13 year old non-clinical children. British Journal of Clinical Psychology 43 (4):437-448

31. Muris P, Meesters C, van den Berg F (2003) The Strengths and Difficulties Questionnaire (SDQ). Further evidence for its reliability and validity in a community sample of Dutch children and adolescents. European Child and Adolescent Psychiatry 12 (1):1-8

32. Niclasen J, Skovgaard AM, Andersen AM, Sømhovd MJ, Obel C (2013) A confirmatory approach to examining the factor structure of the Strengths and Difficulties Questionnaire (SDQ): a large scale cohort study. Journal of Abnormal Child Psychology 41 (3):355-365

33. Niclasen J, Teasdale TW, Andersen AM, Skovgaard AM, Elberling H, Obel C (2012) Psychometric properties of the Danish Strength and Difficulties Questionnaire: the SDQ assessed for more than 70,000 raters in four different cohorts. PLoS One 7 (2):e32025. doi:10.1371/journal.pone.0032025

34. Palmieri PA, Smith GC (2007) Examining the structural validity of the Strengths and Difficulties Questionnaire (SDQ) in a U.S. saple of custodial grandmothers. Psychological Assessment 19 (2):189-198

35. 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 (1):43-48

36. Perera S, Thalagala E, Chandrarathna SH, Agampodi TC, Nugegoda DB, Agampodi SB (2013) Factor structure and normative data of the Sinhalese version of self reported Strength and Difficulties Questionnaire (SDQ) for adolescents. Ceylon Medical Journal 58 (2):66-71

37. Poulou MS (2013) Emotionality and Social Behaviour. Hellenic Journal of Psychology 10:47-60

38. Richter J, Sagatun A, Heyerdahl S, Oppedal B, Rotsamb 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 (9):1002-1011

39. Rodriguez-Hernandez PJ, Betancort M, Ramirez-Santana GM, Garcia R, Sanz-Alvarez EJ, De las Cuevas-Castresana C (2012) Psychometric properties of the parent and teacher versions of the Strenghts and Difficulties Questionnaire (SDQ) in a Spanish sample. International Journal of Clinical and Health Psychology 12 (2):265-279

40. Ronning JA, Handegaard BH, Sourander A, Lorch W-T (2004) The Strengths and Difficulties self-report Questionnaire as a screening instrument in Norwegian community samples. European Child and Adolescent Psychiatry 13 (2):73-82

41. Rothenberger A, Becker A, Erhart M, Wille N, Ravens-Sieberer U, group Bs (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 and Adolescent Psychiatry 17 (Suppl.1):99-105

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

43. Ruchkin V, Koposov R, Swab-Stone M (2007) The Strengths and Difficulties Questionnaire: scale validation with Russian adolescents. Journal of Clinical Psychology 63 (9):861-869

44. Ruchkin V, Koposov R, Vermeiren R, Schwab-Stone M (2012) The Strength and Difficulties Questionnaire: Russian validation of the teacher version and comparison of teacher and student reports. Journal of Adolescence 35 (1):87-96

45. Sanne B, Torsheim T, Heiervang E, Stormark KM (2009) The Strengths and Difficulties Questionnaire in the Bergen Child Study: a conceptually and methodically motivated structural analysis. Psychological Assessment 21 (3):352-364

46. Smedje H, Broman J-E, Hetta J, von Knorring A-L (1999) Psychometric properties of a Swedish version of the "Strengths and Difficulties Questionnaire". European Child and Adolescent Psychiatry 8 (2):63-70

47. Stone LL, Otten R, Engels RCME, Vermulst AA, Janssens JMAM (2010) Psychometric properties of the parent and teacher versions of the Strengths and Difficulties Questionnaire for 4- to 12-years-olds: a review. Clinical Child and Family Psychology Review 13 (3):254-274

48. Tobia V, Grabriele MA, Marzocchi GM (2013) The Italian version of the Strengths and Difficulties Questionnaire (SDQ) - Teacher. Psychometric properties. Journal of Psychoeducational Assessment 31 (5):493-505

49. van de Looij-Jansen PM, Goedhart AW, 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? The British Journal of Clinical Psychology 50 (2):127-144

50. van Leeuwen K, Meerschaert T, Bosmans G, De Medts L, Braet C (2006) The Strengths and Difficulties Questionnaire in a community sample of young children in Flanders. European Journal of Psychological Assessment 22 (3):189-197

51. van Roy B, Veenstra M, Clench-Aas J (2008) Construct validity of the five-factor Strengths and Difficulties Questionnaire (SDQ) in pre-, early-, and late adolescence. Journal of Child Psychology and Psychiatry 49 (12):1304-1312

52. Woerner W, Becker A, Rothenberger A (2004) Normative data and scale properties of the German parent SDQ. European Child and Adolescent Psychiatry 13 (suppl.2):II/3-II/10

53. Yao S, Zhang C, Zhu X, Jing X, McWhinnie CM, Abela JRZ (2009) Measuring adolescent psychopathology: psychometric properties of the self-report Strengths and Difficulties Questionnaire in a sample of Chinese adolescents. Journal of Adolescent Health 45 (1):55-62

54. Zwirs BW, Burger H, Schulpen TW, Wiznitzer M, Fedder H, Buitelaar JK (2007) Prevalence of psychiatric disorders among children of different ethnic origin. Journal of Abnormal Child Psychology 35 (4):556-566