

Investigation of a bifactor model of the Strengths and Difficulties Questionnaire

Hervé Caci · Alexandre J. S. Morin · Antoine Tran

Received: 29 September 2014 / Accepted: 11 January 2015 / Published online: 29 January 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract 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 clear. 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. 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. 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 modelling · Teachers · Externalizing disorders · Internalizing disorders · Prosocial behaviour · SDQ · Measurement invariance

Introduction

Typically, behavioural 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 instruments 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

Electronic supplementary material The online version of this article (doi:10.1007/s00787-015-0679-3) contains supplementary material, which is available to authorized users.

H. Caci (✉) · A. Tran
Hôpitaux Pédiatrique de Nice, CHU Lenval, 57, avenue de la
Californie, 06200 Nice, France
e-mail: caci.h@pediatrie-chulenal-nice.fr

A. J. S. Morin
Institute for Positive Psychology and Education, Australian
Catholic University, Strathfield, Australia

“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 a higher-order ID factor [8]. Although these alternative structures make sense and each has 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 [12] 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 the 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, represent 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 is 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) [25, 26]. 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 $d = 0.39$), Hyperactivity (Cohen's $d = 0.60$), and Peer Relationships (Cohen's $d = 0.17$) factors, and lower levels on the S/PB (Cohen's $d = 0.56$) factor. Similarly, younger children are known to present higher levels on the Emotional (Cohen's $d = 0.11$), Hyperactivity (Cohen's $d = 0.15$) and S/PB (Cohen's $d = 0.08$) 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 mean differences as a function of gender, after having established the measurement invariance of the model [27]. 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 and 18 years old [19, 20, 28]. Overall, 262 teachers participated in the study (mean age = 43.9; SD = 8.6; range = 24–61); 47 were males (17.94 %). Each was asked to rate 2–4 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 = 0.29) 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 = 0.36) 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 [29] 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 answer categories [30–34]. 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 [12] 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 an 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) Disorder factors, correlated to an 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 an 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 an 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, 35]. The sequence of tests is as follows: (a) configural invariance, (b) metric/weak invariance (invariance of the factor loadings); (c) scalar/strong invariance (invariance of the factor loadings and thresholds); (d) strict invariance (invariance of the factor loadings, thresholds and uniquenesses), (e) invariance of the latent variances–covariances (invariance of the factor loadings, thresholds, uniquenesses and variances–covariances), and (f) 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 [36, 37]. Values greater than 0.95 for CFI and TLI are

considered to be indicative of adequate model fit. Values smaller than 0.08 or 0.06 for the RMSEA support, respectively, acceptable and excellent model fit. Chi-square differences tests were computed using MPlus DIFFTEST function $MD\Delta\chi^2$; [38, 39], with the significance level to identify non-invariance fixed at 0.01 to take into account the overall number of $MD\Delta\chi^2$ tests performed [40–42]. 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 0.01 or less and a RMSEA augmentation of 0.015 or less between a model and the preceding model in the invariance sequence indicate that the invariance hypothesis should not be rejected [43, 44].

Scale score reliability for the estimated factor is estimated based on omega (ω) [45] 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 [46]. 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 ($\lambda = 0.413$ – 0.951 ; $M = 0.746$) except item 23 ($\lambda = 0.248$). All scale score reliability coefficients proved satisfactory ($\omega = 0.758$ – 0.914). However, the correlations between the Conduct Problems and Hyperactivity-Inattention factors ($r = 0.748$) and between the Emotional Symptoms and Peer Problems ($r = 0.513$) 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 ($\Delta CFI = -0.002$; $\Delta TLI = -0.002$; $\Delta RMSEA = +0.001$). 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

Table 1 Fit Indices for the alternative measurement models

Model	χ^2	<i>df</i>	CFI	TLI	RMSEA	RMSEA 90 % CI	$\Delta\chi^2$	Δdf	ΔCFI	ΔTLI	$\Delta RMSEA$
M1. 1-Factor model (Global)	3,247.933*	275	0.766	0.744	0.110	0.107–0.114					
M2. 2-Factor model (S & D)	2,795.792*	274	0.801	0.782	0.102	0.098–0.105					
M3. 3-Factor model (S, E, I)	1,981.337*	272	0.865	0.851	0.084	0.081–0.088					
M4. Alternative 3-factor model	1,489.309*	249	0.903	0.892	0.075	0.071–0.078					
M5. 5-factor model	1,349.631*	265	0.914	0.903	0.068	0.064–0.071					
M6. Higher-order (S & D)	1,591.795*	270	0.896	0.884	0.074	0.071–0.078					
M7. Higher-order (S, E, I)	1,387.147*	268	0.912	0.901	0.069	0.065–0.072					
M8. Bifactor (S & D)	1,310.780*	254	0.917	0.902	0.068	0.065–0.072					
M9. Bifactor (S, E, I)	1,080.672*	252	0.935	0.922	0.061	0.057–0.065					
<i>Measurement invariance (M9) by gender groups</i>											
MG1. Configural invariance	1,268.703*	504	0.932	0.919	0.058	0.054–0.062	–	–	–	–	–
MG2. Metric/weak invariance	1,305.754*	542	0.937	0.930	0.056	0.052–0.060	64.290*	38	0.005	0.011	–0.002
MG3. Scalar/strong invariance	1,239.423*	560	0.940	0.935	0.052	0.048–0.056	58.570*	56	0.003	0.005	–0.004
MG4. Strict invariance	1,225.224*	585	0.943	0.942	0.050	0.046–0.054	26.505	25	0.003	0.007	–0.002
MG5. Latent variances-covari- ances invariance	1,199.467*	592	0.946	0.945	0.048	0.044–0.052	11.852	7	0.003	0.003	–0.002
MG6. Latent means invariance	1,392.166*	599	0.929	0.929	0.055	0.051–0.058	110.909*	7	–0.017	–0.016	0.007
<i>Measurement invariance (M9) by school level</i>											
MS1. Configural invariance	1,579.005*	756	0.927	0.913	0.060	0.056–0.065	–	–	–	–	–
MS2. Metric/weak invariance	1,596.525*	832	0.932	0.926	0.055	0.051–0.060	118.118*	76	0.005	0.013	–0.005
MS3. Scalar/strong invariance	1,611.085*	868	0.934	0.931	0.054	0.049–0.058	45.037	36	0.002	0.005	–0.001
MS4. Strict invariance	1,638.032*	904	0.935	0.935	0.052	0.048–0.056	76.210*	36	0.001	0.004	–0.002
MS5. Latent variances-covari- ances invariance	1,617.358*	918	0.938	0.939	0.051	0.046–0.055	23.650	14	0.003	0.004	–0.001
MS6. Latent means invariance	1,628.433*	932	0.938	0.940	0.050	0.046–0.054	28.308*	14	0.000	0.001	–0.001

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 the resulting CFI values can be non-monotonic with model complexity

S Strengths, *D* Difficulties, *E* Externalizing disorders, *I* Internalizing disorders, 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, *MD* $\Delta\chi^2$ Chi square difference test calculated with the Mplus DIFFTEST function for the robust weighted least square estimator (WLSMV)

* *p* < 0.01

M9 proved fully acceptable and substantially better than the fit of alternative models M5, M7 and M8 according to the goodness-of-fit indices ($\Delta CFI = +0.018$ – $+0.023$; $\Delta TLI = +0.019$ – 0.021 ; $\Delta RMSEA = -0.007$ – -0.008), although the confidence interval of the RMSEA suggests that these differences may not be fully significant.

Model M9 was thus retained (Fig. 1), 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 ($\lambda = 0.679$ – 0.830 and $\lambda = 0.467$ – 0.791 , respectively), and present a satisfactorily high level of scale score reliability ($\omega = 0.883$ and $\omega = 0.926$, 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 ($\lambda = -0.119$ – 0.492) and Hyperactivity-Inattention ($\lambda = -0.161$ – 0.758) 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, $\lambda = 0.492$; steals from home, school or elsewhere, $\lambda = 0.439$). As a result, the scale score reliability of this S-factor remains quite low ($\omega = 0.544$). 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, $\lambda = 0.622$; constantly fidgeting or squirming, $\lambda = 0.758$), rather than Inattention. Given the magnitude of these two specific loadings, the scale score reliability of this S-factor remains generally satisfactory ($\omega = 0.705$).

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)

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		

All coefficients are significant ($p \leq 0.05$)

4 items with $\lambda > 0.500$ 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 $0.200 \leq \lambda < 0.500$ 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 ($\omega = 0.804$). The S-factor reflecting Emotional Symptoms going over and above this generic presence of ID is also well defined through high-factor loadings ($\lambda = 0.468$ – 0.780), and a satisfactory level of scale score reliability ($\omega = 0.832$). 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, $\lambda = 0.570$; gets along better with adults than with other children, $\lambda = 0.707$) than items reflecting peer rejection ($\lambda = 0.123$ – 0.313) which present stronger relations with the ID G-factor—resulting in a lower S-factor scale score reliability estimate ($\omega = 0.650$).

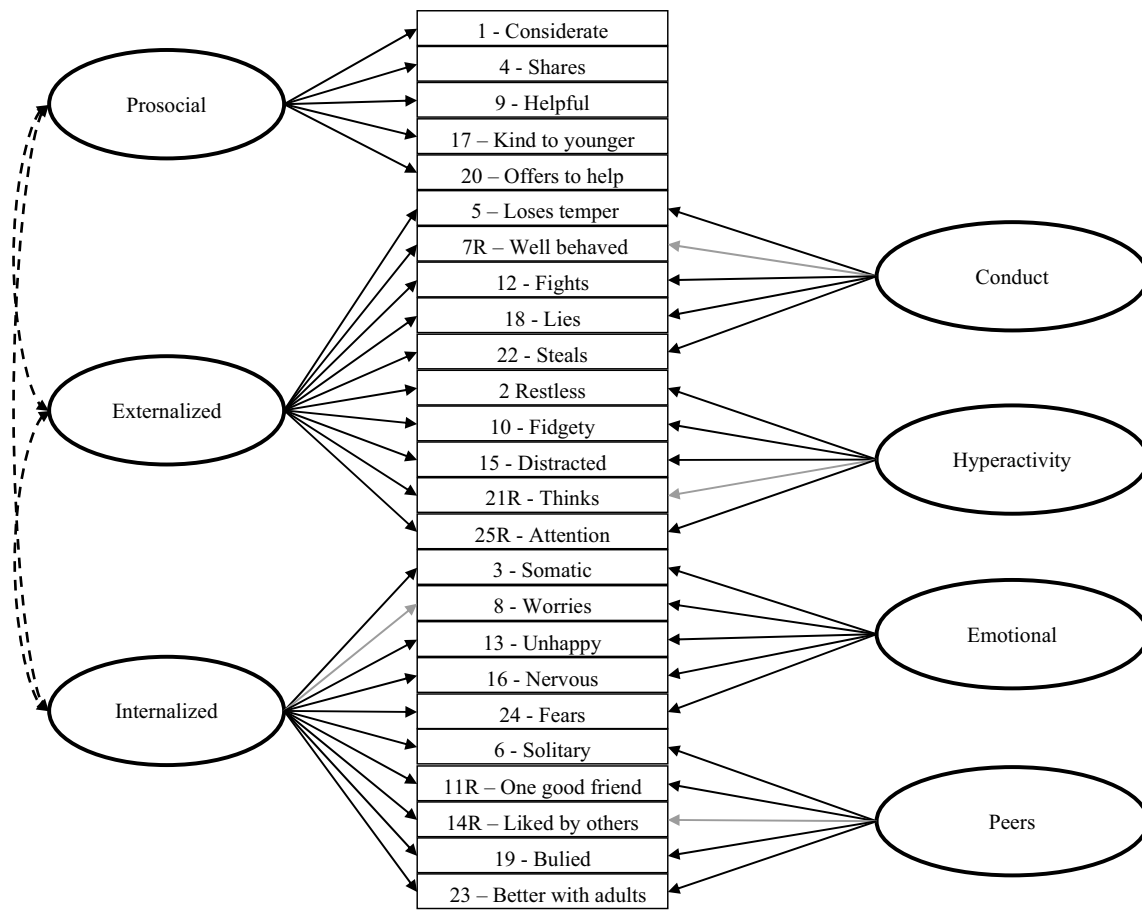


Fig. 1 Final retained bifactor Model M9: loadings are represented as *solid lines with an arrow* (non-significant loadings are in *grey*) and factor correlations are represented as *broken lines with arrows* at both ends (see numerical values in Table 3)

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 $\Delta RMSEA$ exceeded the recommended cutoffs. 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 mean differences across school levels ($\Delta CFI = 0.000$; $\Delta TLI = +0.001$; $\Delta RMSEA = -0.001$), and the presence of latent mean differences across genders ($\Delta CFI = -0.017$; $\Delta TLI = -0.016$; $\Delta RMSEA = +0.007$). 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, $p < 0.05$), the ID G-factor (0.252, $p < 0.05$) and on the Hyperactivity/Inattention S-factor (0.287, $p \leq 0.05$), but lower latent means on the Strength/Prosocial Behaviour factor (-0.467 , $p < 0.05$) and the Emotional Problems

S-factor (-0.231 , $p < 0.05$). 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

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		<i>-0.119</i>				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			<i>-0.118</i>			0.791		0.361
25-R	Attention			<i>-0.161</i>			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	Factor 1	Factor 1	Factor 1	Factor 1	Factor 1	
	G-Factor 1								<i>-0.730**</i>
	G-Factor 2								<i>-0.635**</i>
									<i>0.561**</i>

Non-significant coefficients are marked in italics

All other coefficients are significant ($p \leq 0.05$)

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 an 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 [47]. Rather, the specific patterns of significant versus non-significant loadings helped us to refine the interpretation of the G- and S-factors.

Parameter 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 ($\omega = 0.804\text{--}0.926$), 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 the items, and presents satisfactory scale score reliability ($\omega = 0.832$), 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 [48–50]. 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 ($\omega = 0.544$) 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 to 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 [51, 52].

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 [53, 54]. 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 ($\lambda = 0.707$). The level of specificity related to these S-factors (Hyperactivity-Inattention and Peer Problems) is sufficient to provide satisfactory scale score reliability estimates that fully justify their use to complement ratings on the G-factors to assess hyperactivity ($\omega = 0.705$) and preference for solitude ($\omega = 0.650$), 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 level differences could be observed as a function of school levels. Conversely, but also supporting previous studies based either on the SDQ [27] 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 of 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 [55] due to their reliance on distinct frames of references in their interaction with the child being rated. In addition, the generalizability of the present results to other versions of SDQ and other linguistic groups should also be more thoroughly investigated.

Acknowledgments The ChiP-ARD study was funded 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.

Conflict of interest The authors have no conflict of interests to declare.

References

1. American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders, 5th edn. American Psychiatric Press, Washington
2. Preisig M, Waeber G, Vollenweider P, Bovet P, Rothen S, Vandelour C, Guex P, Middleton L, Waterworth D, Mooser V, Tozzi F, Muglia P (2009) The PsyCoLaus study: methodology and characteristics of the sample 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. *Psychol Med* 35(2):245–256
4. Goodman R (1997) The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol Psychiatry* 38(5):581–586
5. Rutter M (1967) A children's behaviour questionnaire for completion by teachers: preliminary findings. *J Child Psychol 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. *J Child Psychol Psychiatry* 35(8):1483–1494
7. Goodman R (2001) Psychometric properties of the strengths and difficulties questionnaire. *J Am Acad Child Adolesc 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. *J Abnorm Child Psychol* 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, 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, 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. *J Exp Educ*
12. Dickey WC, Blumberg SJ (2004) Revisiting the factor structure of the strengths and difficulties questionnaire: United States, 2001. *J Am Acad Child Adolesc Psychiatry* 43(9):1159–1167
13. Brunner M, Nagy G, Wilhelm O (2012) A tutorial on hierarchically structured constructs. *J Pers* 80(4):796–846. 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. *Multivar Behav Res* 23(1):51–67
15. Chen FF, West SG, Sousa KH (2006) A comparison of bifactor and second-order models of quality of life. *Multivar Behav Res* 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. *J Pers Assess* 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. *Multivar Behav Res* 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. *J Atten Disord*
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. *J Atten Disord*
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 strengths and difficulties questionnaire. *Eur J Psychol Assess* 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 and Francis, New York
25. Niclasen J, Skovgaard AM, Andersen AM, Somhøvd MJ, Obel C (2013) A confirmatory approach to examining the factor structure of the Strengths and Difficulties Questionnaire (SDQ): a large scale cohort study. *J Abnorm Child Psychol* 41(3):355–365
26. Niclasen J, Teasdale TW, Andersen AM, Skovgaard MJ, Elberling H, Obel C (2012) Psychometric properties of the Danish Strengths and Difficulties Questionnaire: the SDQ assessed for more than 70,000 raters in four different cohorts. *PLoS One* 7(2):e32025
27. d'Acremont M, Van der Linden M (2008) Confirmatory factor analysis of the strengths and difficulties questionnaire in a community of French-speaking adolescents. *Eur J Psychol Assess* 24(1):1–8
28. Caci H, Morin AJS, Tran A (2014) Prevalence and correlates of attention deficit/hyperactivity disorder in adults from a French community sample. *J Nerv Mental Dis* 202(4):324–332
29. Muthén LK, Muthén BO (2012) *Mplus User's Guide*, 7th edn. Muthén and Muthén, Los Angeles
30. Beauducél A, Herzberg PY (2006) On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Struct Equ Model* 13(2):186–203
31. 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, pp 269–314
32. Flora DB, Curran PJ (2004) An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychol Methods* 9(4):466–491
33. Forero CG, Maydeu-Olivares A, Gallardo-Pujol D (2009) Factor analysis with ordinal indicators: a Monte Carlo study comparing DWLS and ULS estimation. *Struct Equ Model* 16(4):625–641
34. 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)
35. Millsap RE, Tein JY (2004) Assessing factorial invariance in ordered-categorical measures. *Multivar Behav Res* 39(3): 479–515
36. Hu LT, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model* 6(1):1–55
37. Yu CY (2002) Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes. University of California, Los Angeles
38. Asparouhov T, Muthén BO (2006) *Robust Chi square difference testing with mean and variance adjusted test statistics*. Los Angeles, Muthén and Muthén; <http://www.statmodel.com/examples/webnote.shtml> - web10
39. Muthén BO (2004) *Mplus Technical Appendices*. Los Angeles, Muthén and Muthén; <http://www.statmodel.com/techappen.shtml>
40. Bollen KA (1989) *Structural equations with latent variables*. Wiley, New York
41. Morin AJS, Madore I, Morizot J, Boudrias J-S, Tremblay M (2009) The workplace affective commitment multidimensional questionnaire: factor structure and measurement invariance. *Int J Psychol Res* 4(3–4):307–344
42. Rensvold RB, Cheung GW (1998) Testing measurement model for factorial invariance: a systematic approach. *Educ Psychol Meas* 58(6):1017–1034
43. Chen FF (2007) Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct Equ Model* 14(3):464–504
44. Cheung GW, Rensvold RB (2002) Evaluating goodness-of fit indexes for testing measurement invariance. *Struct Equ Model* 9(2):233–255
45. McDonald RP (1970) The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *Br J Math Stat Psychol* 23(1):1–21
46. 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
47. Morin AJS, Arens AK, Marsh HW (In press) A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Struct Equ Model*
48. Le Corff Y, Toupin J (Ahead of print) Overt versus covert conduct disorder symptoms and the prospective prediction of antisocial personality disorder. *J Personal Dis*. doi:10.1521/pedi_2013_27_074
49. 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. *J Am Acad Child Adolesc Psychiatry* 41(3):253–261
50. Loeber R, Schmaling KB (1985) Empirical evidence for overt and covert patterns of antisocial conduct problems: a metaanalysis. *J Abnorm Child Psychol* 13(2):337–353
51. 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. *J Abnorm Child Psychol* 33(5):537–550
52. 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
53. Rubin KH, Copland RJ, Bowker JC (2009) Social withdrawal in childhood. *Annu Rev Psychol* 60:141–171
54. Wang JM, Rubina KH, Laursen B, Booth-LaForce C, Rose-Krasnord L (2013) Preference-for-solitude and adjustment difficulties in early and late adolescence. *J Clin Child Adolesc Psychol* 42(6):834–842
55. Merwood A, Grevén CU, Price TS, Fijdsdijk F, Kuntsi J, McLoughlin G, Larsson H, Asherson PJ (2013) Different heritabilities but shared etiological influences for parent, teacher and self-ratings of ADHD symptoms: an adolescent twin study. *Psychol Med* 43(9):1973–1984