

The Strengths and Difficulties Questionnaire (SDQ) parent-report: bifactor structure and method effects in a community-based sample

Örnólfur Thorlacius & Auðun Valborgarson

To cite this article: Örnólfur Thorlacius & Auðun Valborgarson (19 Apr 2026): The Strengths and Difficulties Questionnaire (SDQ) parent-report: bifactor structure and method effects in a community-based sample, Nordic Psychology, DOI: [10.1080/19012276.2026.2655128](https://doi.org/10.1080/19012276.2026.2655128)

To link to this article: <https://doi.org/10.1080/19012276.2026.2655128>



© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



View supplementary material [↗](#)



Published online: 19 Apr 2026.



Submit your article to this journal [↗](#)



Article views: 531



View related articles [↗](#)



View Crossmark data [↗](#)

The Strengths and Difficulties Questionnaire (SDQ) parent-report: bifactor structure and method effects in a community-based sample

ÖRNÓLFUR THORLACIUS¹ & AUÐUN VALBORGARSON²

Correspondence address: Örnólfur Thorlacius Department of Psychology, University of Iceland, Reykjavík, Iceland.
Email: ort@hi.is

Abstract

The Strengths and Difficulties Questionnaire (SDQ) is a widely used screening measure for child mental health, yet findings regarding its latent structure have been inconsistent. The present study examined the structural validity of the SDQ parent-report in a large community sample of mothers of Icelandic children aged 6–12 years ($N = 2332$). Competing measurement models were evaluated, including the original five-factor structure, second-order and bifactor models, estimated with and without a method factor capturing covariance among positively worded items. Model comparisons indicated that a bifactor model including two broad general factors—Externalizing and Internalizing problems—together with a Prosocial Behavior factor and an orthogonal wording-related method factor provided the most interpretable representation. The Externalizing ($\omega_H = 0.69$; $ECV = 0.65$) and Internalizing ($\omega_H = 0.58$; $ECV = 0.55$) factors accounted for a substantial proportion of reliable variance among the symptom-based items, whereas the specific SDQ factors, particularly Conduct Problems and Peer Problems, contributed relatively limited variance beyond these broader factors. The wording method factor accounted for a notable share of covariance among positively worded items, particularly those from the Prosocial Behavior domain, consistent with structured wording-related variance. Sensitivity analyses using alternative specifications, including exploratory structural equation modeling and a random-intercept factor approach, yielded converging results. The findings highlight the influence of wording-related covariance and skewed and restricted response distributions, underscoring the importance of accounting for these sources of variance when interpreting SDQ scores. Broad Internalizing and Externalizing scores may therefore provide more stable indicators of child mental health difficulties in non-clinical populations.

Keywords: Strengths and Difficulties Questionnaire, bifactor analysis, wording effects, parent ratings, child psychopathology

¹Department of Psychology, University of Iceland, Reykjavík, Iceland;

²Department of Psychology, Reykjavík University, Reykjavík, Iceland

Supplemental data for this article can be accessed online at <https://doi.org/10.1080/19012276.2026.2655128>.

Introduction

The Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) is a widely used screening tool for assessing mental health problems in children and youth. The measure was designed to reflect a five-factor structure comprising the following scales: Emotional problems, Conduct problems, Hyperactivity/Inattention, Peer Problems, and Prosocial Behavior (Goodman, 1997). Research examining the latent structure of the SDQ, however, has produced inconsistent findings across samples and analytic approaches, raising questions about the instrument's dimensional structure (Caci et al., 2015; Goodman et al., 2010; McCrory & Layte, 2012; Stone et al., 2010; Van Leeuwen et al., 2006; Vugteveen et al., 2021). Prior research suggests that covariance among SDQ items may arise from multiple sources beyond the simple five-factor representation originally proposed. In particular, shared variance among items may reflect broader psychopathology dimensions such as internalizing and externalizing factors (e.g., Goodman et al., 2010; Kóbor et al., 2013), or multidimensional items that produce cross-loadings across domains (e.g., Garrido et al., 2020; Ortuño-Sierra et al., 2015). In addition, systematic covariance related to item wording may also influence the SDQ's latent structure (e.g., McCrory & Layte, 2012; Palmieri & Smith, 2007; Van Roy et al., 2008). Each of these possibilities has been discussed in the SDQ literature and has implications for how the questionnaire's latent structure should be modeled and interpreted.

Evidence from confirmatory factor-analytic studies suggests that the covariance among SDQ items may not be fully captured by the five correlated domains alone. Although exploratory factor analysis (EFA) often recovers the original five-factor structure, results from confirmatory factor analysis (CFA) have been less consistent. In a systematic review of more than 50 factor-analytic studies, Caci et al. (2015) concluded that CFA models frequently show inadequate fit unless model modifications are introduced, such as allowing residual correlations among items (e.g., Ortuño-Sierra et al., 2015) or removing problematic items (e.g., Duinhof et al., 2019). These findings suggest that alternative modeling approaches may be required to adequately represent the relationships among SDQ items.

Bifactor modeling has been proposed as a useful framework for understanding the SDQ's multidimensional structure (e.g., Caci et al., 2015; Kóbor et al., 2013). In a traditional bifactor model (e.g., Holzinger & Swineford, 1937; Rodriguez et al., 2016), each item loads onto a general factor representing variance common across all items, as well as a specific factor capturing residual variance associated with item groupings (e.g., Emotional problems, Conduct problems). This approach helps disentangle general from domain-specific symptom variance and has been recommended for evaluating multidimensional psychological scales (e.g., Rodriguez et al., 2016), consistent with hierarchical models of psychopathology that distinguish broad and specific dimensions of symptoms (e.g., Krueger & Markon, 2006; Lahey et al., 2012). However, bifactor models have also been criticized for potentially overfitting data, yielding poorly defined or unreliable specific factors, and complicating substantive interpretation when general factors account for a large proportion of the common variance (e.g., Bonifay et al., 2017; Reise et al., 2013).

Few studies have applied bifactor modeling to test the structural validity of the SDQ. Kóbor et al. (2013) found that a general bifactor model fit best in teacher- and parent-reported data, retaining the five SDQ subscales as specific factors alongside a broad general

problems factor. Similarly, Skoczeń et al. (2018) provided support for a general bifactor structure for the self-reported SDQ in a Polish adolescent sample, highlighting a strong general difficulties factor. In contrast, Bell et al. (2024) favored the original five-factor SDQ structure over a traditional general bifactor model in a large U.S. teacher-reported sample, primarily due to low scale score reliability of the specific factors in the bifactor model and the resulting concerns about subscale interpretability.

To date, only one study has explicitly tested a bifactor structure aligned with the internalizing–externalizing distinction, a well-established framework in developmental psychopathology used to organize common mental health problems in childhood and adolescence (e.g., Krueger & Markon, 2006; Lahey et al., 2012). Caci et al. (2015) applied an internalizing–externalizing bifactor model to teacher-reported SDQ data in a French community sample. In this specification, Emotional Problems and Peer Problems items loaded on a general Internalizing factor alongside their domain-specific factors, whereas Conduct Problems and Hyperactivity/Inattention items loaded on a general Externalizing factor and their respective domain-specific factors. Caci et al. found that the internalizing–externalizing bifactor model provided the best fit, indicating that covariance among the SDQ items may partly reflect broader symptom dimensions consistent with hierarchical models of child psychopathology.

A further psychometric issue, and potential source of covariance, concerns the SDQ's use of both positively and negatively worded items. The inclusion of differently phrased items can introduce wording-related method effects that compromise factorial validity by adding systematic covariance unrelated to the constructs being assessed (McCrary & Layte, 2012; Palmieri & Smith, 2007; Van Roy et al., 2008). Of the 25 SDQ items, 10 are positively worded (e.g., *Considerate of other people's feelings*) while the remaining 15 are negatively worded (e.g., *Often fights with other children or bullies them*). Importantly, five of the positively worded items in the SDQ are designed to load on symptom or difficulty factors (e.g., Emotional Symptoms, Conduct Problems), despite their positive phrasing. Such wording differences may cause these items to correlate due to shared response tendencies rather than substantive similarity.

To address this, several studies have incorporated a method factor for positively worded items in their SDQ models (e.g., Palmieri & Smith, 2007; Van Roy et al., 2008). This approach accounts for shared variance among the items and has consistently improved model fit. Alternative strategies include allowing cross-loadings or residual correlations among similarly worded items (e.g., van de Looij-Jansen et al., 2011) or removing problematic items altogether (e.g., Duinhof et al., 2019; Kankaanpää et al., 2023). For example, Van Roy et al. (2008) highlighted the impact of positively worded items on model fit and invariance in the SDQ self-report. Collectively, these findings underscore the substantial influence of item wording on the SDQ's factor structure and the importance of accounting for this source of method variance when modeling and interpreting SDQ scores.

The current study

The aim of this study was to examine the latent structure of the SDQ parent-report in a large community sample of Icelandic mothers of children aged 6–12 years. Tested models included the original five-factor SDQ structure as well as second-order and bifactor models reflecting the well-established internalizing–externalizing distinction in child

psychopathology. Although several studies have applied bifactor modeling to the SDQ (e.g., Bell et al., 2024; Caci et al., 2015; Kóbor et al., 2013; Skoczeń et al., 2018), most have focused on self- or teacher-reports, explored a limited number of alternative structures, and produced inconsistent findings.

The present study therefore compared alternative structural models for the SDQ parent-report—including second-order and bifactor specifications reflecting the internalizing–externalizing distinction—in a community-based sample, while accounting for wording-related covariance among positively worded items. In addition, sensitivity analyses were conducted to evaluate whether conclusions regarding the retained model were robust to key modeling assumptions. Specifically, an exploratory structural equation modeling (ESEM) model with target rotation was estimated to relax the CFA assumption that items load only on their designated factors, with non-target cross-loadings fixed to zero (e.g., Garrido et al., 2020). Moreover, a random-intercept item factor analysis (RIIFA; Maydeu-Olivares & Coffman, 2006; Aichholzer, 2014) model was estimated to examine whether covariance among items could be accounted for by a general response-style factor.

Method

Participants and procedure

Data were drawn from two published studies (Thorlacius & Gudmundsson, 2015, 2019) and an unpublished dataset (Valborgarson, 2016), all collected in Iceland during 2014–2018, following the same study protocol. Participants were recruited *via* schools and invited to complete an online survey after approval was obtained from the university science and ethics board, local education offices, and schools. All mothers of children in the target age range were eligible to participate, and no additional exclusion criteria were applied.

The combined data sample consisted of maternal ratings of 2332 children aged 6–12 years ($M = 9.3$, $SD = 1.9$; 51.5% boys). Most mothers were Icelandic (97.9%), lived in the greater capital area of Reykjavík (45.6%), and had a first university degree (e.g., BA, BS or BEd degree; 35.2%), or a master's or doctoral degree (25.8%). Information on marital status was only available for two-thirds of participants, most of which reported their status as married ($\approx 62\%$) or cohabiting ($\approx 26\%$).

Because the present study pooled data across three samples, potential differences in sample composition and factor structure were evaluated across datasets. Table S1 presents sample characteristics by dataset, indicating minor demographic differences (e.g., maternal education and geographic distribution), whereas age and sex distributions were broadly comparable. Configural invariance was tested across datasets, sex, and age group (6–8, 9–10, and 11–12 years) using multi-group CFA. The configural model across datasets showed acceptable fit, $\chi^2(795) = 2640.09$, CFI = 0.968, TLI = 0.964, RMSEA = 0.056, indicating a similar loading pattern across groups. Comparable fit was observed across sex and age groups (see Table S2), supporting estimation of the structural models in the combined data sample.

The Strengths and Difficulties Questionnaire

The SDQ (Goodman, 1997) consists of 25 items, each rated as being *not true* (0), *somewhat true* (1), or *certainly true* (2). The items are grouped into four symptom-based/difficulty

scales (Emotional Problems, Peer Problems, Conduct Problems, and Hyperactivity/Inattention) and one strength-based scale (Prosocial Behavior). Five items from the difficulty scales (items 7, 11, 14, 21, 25) are positively worded and should thus be reverse scored (Goodman, 1997). According to the SDQ manual, a Total Difficulties score can be calculated by summing the scores of the 20 difficulties items (i.e., all items except those belonging to the Prosocial Behavior scale). The Icelandic version of the SDQ parent-report was used in the study, available at www.sdqinfo.org.

Statistical analysis

All statistical analyses were performed in R (version 4.2.1; R Core Team, 2022). Item analysis was conducted with the *psych* package (version 2.2.3, Revelle, 2022) and structural modeling was performed using the *Lavaan* package (version 0.6-15, Rosseel, 2023). Because missing values were minimal (<5% per item), imputation was deemed appropriate (Tabachnick & Fidell, 2019), and missing data were imputed with the *mice* package (v3.15.0; Van Buuren & Groothuis-Oudshoorn, 2011), from which a single completed dataset was retained for subsequent analyses.

In the structural analyses, the diagonal weighted least squares (DWLS) estimator was used, appropriate for ordinal data (DiStefano & Morgan, 2014) and model fit was evaluated according to commonly accepted indices and criteria: Root mean square error of approximation (RMSEA) ≤ 0.06 , comparative fit index (CFI) ≥ 0.95 , and Tucker-Lewis index (TLI) ≥ 0.95 (Hu & Bentler, 1999). Nested models were compared using scaled chi-square difference ($\Delta\chi^2$) tests appropriate for DWLS estimation. Given evidence that global fit indices may behave differently across hierarchical and bifactor specifications, model comparisons were interpreted cautiously and in conjunction with parameter estimates and scale score reliability indices (Greene et al., 2019).

The following measurement models were examined: (1) A correlated five-factor CFA model corresponding to the original SDQ structure; (2–3) Extensions of the five-factor model including a wording factor defined by either five positively worded difficulty items or all ten positively worded items; (4–5) A second-order model specifying a single Total Difficulties factor (over the four difficulty domains) with a separate Prosocial Behavior factor, estimated with and without a wording factor; (6–7) A second-order model specifying two second-order factors—Externalizing problems (Conduct Problems and Hyperactivity/Inattention) and Internalizing problems (Emotional Symptoms and Peer Problems)—alongside a Prosocial Behavior factor, estimated with and without a wording factor; (8–9) A bifactor model including a general Total Difficulties factor and domain-specific factors, with a separate Prosocial Behavior factor, estimated with and without a wording factor; (10–11) A bifactor model including two general factors (Externalizing and Internalizing problems) and domain-specific factors, together with a Prosocial Behavior factor, estimated with and without a wording factor. In all models including a wording factor, this factor was specified as orthogonal to the substantive factors. In the second-order and bifactor models, domain-specific factors were specified as orthogonal, whereas correlations were freely estimated among the broad factors (Externalizing and Internalizing) and between these factors and Prosocial Behavior.

Scale score reliability was estimated using model-based omega coefficients calculated with the *BifactorIndicesCalculator* (version 0.2.2; Dueber, 2022), including hierarchical omega

(ω H) and omega subscale (ω HS). ω H indexes the proportion of reliable variance attributable to the general factor, whereas ω HS reflects reliable variance in specific factors after controlling for the general factor. Higher ω H values indicate greater general factor saturation (Reise, 2012; Rodriguez et al., 2016), and values around 0.50 or higher have sometimes been taken to indicate substantial general factor influence, although such heuristics should be interpreted cautiously. Explained common variance (ECV) was also computed to characterize the relative contribution of general versus specific factors. Higher ECV values indicate greater dominance of the general factor, whereas lower ECV values in combination with non-trivial ω HS values suggest meaningful specific-factor variance (Reise et al., 2013; Rodriguez et al., 2016).

Additionally, construct replicability was evaluated using Hancock's H coefficient, which reflects the extent to which a latent factor is well-defined by its indicators and likely to be stable across samples (Hancock & Mueller, 2001). Values approaching 0.70 or higher were taken to indicate adequately defined latent constructs, although this criterion was interpreted in conjunction with other model indices.

A sensitivity analysis was conducted using ESEM with target rotation to evaluate whether conclusions regarding the retained model were robust to relaxing the assumption that non-target cross-loadings are fixed to zero in the CFA framework. ESEM permits items to load on multiple factors within a predefined structure while retaining an overall confirmatory orientation (Asparouhov & Muthén, 2009; Marsh et al., 2014). The model was estimated in *lavaan* using the same estimator as in the primary analyses, with target rotation aligning primary loadings with the hypothesized structure while allowing non-target loadings to be freely estimated but targeted toward zero. The analysis examined whether allowing cross-loadings altered the primary loading pattern, parameter estimates, or factor interpretability.

Finally, as an alternative to the wording-factor specification, a RIIFA model was estimated to examine whether item covariance could be accounted for by a general response-style factor. All items were specified to load equally on a random intercept factor representing a general response tendency, specified as orthogonal to the substantive factors (Aichholzer, 2014; Maydeu-Olivares & Coffman, 2006). This model was included as a sensitivity analysis to evaluate whether covariance among items could be accounted for by a general response-style factor rather than by the wording-factor specification.

Results

Preliminary item analysis

Table 1 presents descriptive statistics for the SDQ items, categorized by item valence (negatively vs. positively worded items) and ordered according to level of skewness. Negatively worded difficulty items generally had low mean scores and highly positively skewed distributions. Conduct Problems items reflecting low base-rate behaviors, such as item 22 (*Steals from home, school or elsewhere*), item 12 (*Often fights with other children or bullies them*) and item 18 (*Often lies or cheats*) were endorsed with the highest response option (*Certainly true*) by fewer than 2% of respondents, resulting in pronounced skewness and floor effects.

Positively phrased items, particularly within the Prosocial Behavior scale, were skewed in the opposite direction. Items such as item 17 (*Kind to younger children*), item 9 (*Helpful if someone is hurt, upset or feeling ill*), and item 4 (*Considerate of other people's feelings*) had

Table 1. Descriptive statistics for SDQ items ($N = 2332$).

Item content ^a	Factor	M	SD	Skewness	Kurtosis	Scale response category	
						Low (0–1) ^c	High (2)
Negatively worded items							
22. Steals	Conduct problems	0.05	0.26	5.44	31.30	0.99	0.01
12. Fights	Conduct problems	0.05	0.23	5.03	27.12	0.99	<0.01
18. Lies	Conduct problems	0.18	0.44	2.40	5.17	0.98	0.02
19. Bullied	Peer problems	0.30	0.55	1.69	1.87	0.95	0.05
23. Better with adults	Peer problems	0.34	0.59	1.51	1.22	0.94	0.06
13. Unhappy	Emotional problems	0.35	0.60	1.50	1.14	0.93	0.07
2. Restless	Hyperactivity/inattention	0.41	0.66	1.34	0.50	0.90	0.10
5. Loses temper	Conduct problems	0.42	0.65	1.28	0.41	0.91	0.09
6. Solitary	Peer problems	0.44	0.64	1.28	0.19	0.92	0.08
10. Fidgety	Hyperactivity/inattention	0.47	0.71	1.17	−0.06	0.87	0.13
16. Nervous	Emotional problems	0.48	0.66	1.02	−0.13	0.91	0.09
24. Fears	Emotional problems	0.51	0.68	0.97	−0.31	0.89	0.11
3. Somatic	Emotional problems	0.54	0.71	0.93	−0.47	0.87	0.13
8. Worries	Emotional problems	0.64	0.71	0.65	−0.82	0.86	0.14
15. Distracted		0.80	0.76	0.36	−1.20	0.79	0.21

(Continued)

Table 1. (Continued).

Item content ^a	Factor	M	SD	Skewness	Kurtosis	Scale response category	
						Low (0–1) ^c	High (2)
Positively worded items	Hyperactivity/ inattention						
17. Kind to younger	Prosocial behavior	1.85	0.40	–2.71	6.97	0.02	0.98
11. One good friend ^b	Peer problems	0.20	0.49	2.47	5.27	0.04	0.96
14. Liked by others ^b	Peer problems	0.32	0.53	1.39	0.96	0.03	0.97
9. Helpful	Prosocial behavior	1.68	0.52	–1.27	0.59	0.02	0.98
4. Shares	Prosocial behavior	1.66	0.54	–1.27	0.64	0.03	0.97
1. Considerate	Prosocial behavior	1.59	0.54	–0.80	–0.51	0.02	0.98
7. Obedient ^b	Conduct problems	0.50	0.60	0.78	–0.38	0.06	0.94
20. Offers to help	Prosocial behavior	1.42	0.61	–0.55	–0.61	0.06	0.94
25. Attention ^b	Hyperactivity/ inattention	0.81	0.71	0.29	–0.98	0.17	0.83
21. Thinks ^b	Hyperactivity/ inattention	0.83	0.61	0.12	–0.47	0.12	0.88

Note. SDQ = Strengths and Difficulties Questionnaire; M = Mean; SD = standard deviation.
^a Item labels are abbreviations of the original SDQ items.
^b Reverse-scored item.
^c For clarity, the two lowest response options (Not true and Somewhat true) are displayed together under a single Low (0–1) endorsement category.

means ranging from 1.59 to 1.85, with 97–98% endorsing the highest response option. A similar pattern was observed for reverse-coded difficulty items in the Peer Problems factor (e.g., item 11, *Has at least one good friend*) and the Conduct Problems factor (item 7, *Generally obedient, usually does what adults request*), where 94–97% endorsed the highest category, indicating substantial skewness and ceiling effects.

Model fit evaluation

Table 2 presents fit statistics for all competing models. The correlated five-factor CFA model (Model 1) demonstrated marginal fit (CFI = 0.921, TLI = 0.911, RMSEA = 0.060). Inter-factor correlations indicated substantial overlap among SDQ domains, particularly between Hyperactivity/Inattention and Conduct Problems ($r = 0.73$) and between Prosocial Behavior and Conduct Problems ($r = -0.71$), suggesting considerable shared variance across domains.

Incorporating a wording factor defined by the five positively worded difficulty items (Model 2) yielded a modest improvement in fit. However, a large modification index (MI > 600) for the covariance between the wording factor and Prosocial Behavior indicated substantial strain under the orthogonality constraint, suggesting that wording-related and prosocial variance were not fully separable under this specification. Defining the wording factor using all ten positively worded SDQ items (Model 3) resulted in a substantial improvement in fit (CFI = 0.955, TLI = 0.947, RMSEA = 0.047), with a statistically significant improvement relative to Model 1 ($\Delta\chi^2(10) = 384.99$, $p < .001$). This pattern indicates that wording-related covariance is not limited to reverse-coded difficulty items but extends to positively worded prosocial items. In subsequent models, the wording factor was defined using all ten positively worded items unless otherwise specified.

Second-order models showed comparatively poorer performance. The specification separating Total Difficulties and Prosocial Behavior (Model 4) fit worse than the correlated five-factor model, although inclusion of the wording factor using all ten positively worded SDQ items (Model 5) improved fit ($\Delta\chi^2(10) = 345.88$, $p < .001$). Alternative second-order structures distinguishing Externalizing and Internalizing problems (Models 6 and 7) also showed weaker fit, indicating that hierarchical representations alone did not adequately capture the covariance structure.

In contrast, bifactor models provided improved representation of the data. The bifactor model separating Total Difficulties and Prosocial Behavior (Model 8) showed fit comparable to the correlated five-factor model, and inclusion of the wording factor (Model 9) significantly improved fit ($\Delta\chi^2(10) = 276.20$, $p < .001$). Further improvement was observed for the bifactor model distinguishing Externalizing and Internalizing general factors (Model 10), with additional gains following inclusion of the wording factor (Model 11; $\Delta\chi^2(10) = 320.73$, $p < .001$). This model provided the best overall fit (CFI = 0.968, TLI = 0.960, RMSEA = 0.040) and yielded the clearest substantive interpretation. Subsequent interpretation therefore focuses on Model 11. A schematic representation of the model is shown in [Figure 1](#), and parameter estimates and model-based scale score reliability indices are reported in [Table 3](#).

Retained model structure

General symptom factors

Model 11 revealed two broad general dimensions corresponding to Externalizing and Internalizing problems. Factor-level indices indicated that the Externalizing ($\omega_H = 0.69$,

Table 2. Fit indices for competing measurement models of the SDQ (*N* = 2332).

Model	Model specification	χ^2	df	CFI	TLI	RMSEA (CI 90%)
Correlated five-factor models						
1	CFA five-factor	2512.00	265	0.921	0.911	0.060 (0.058–0.062)
2	Model 1 + wording factor (5 items) ^a	2290.16	260	0.928	0.917	0.058 (0.056–0.060)
3	Model 1 + wording factor (10 items) ^b	1541.91	255	0.955	0.947	0.047 (0.044–0.049)
Second-order models						
4	Second-order (TD + PB)	2869.76	270	0.909	0.898	0.064 (0.062–0.066)
5	Model 4 + wording factor (10 items)	1972.80	260	0.940	0.931	0.053 (0.051–0.055)
6	Second-order (Ext + Int + PB)	4318.29	269	0.858	0.841	0.080 (0.078–0.082)
7	Model 6 + wording factor (10 items)	2822.13	259	0.910	0.896	0.065 (0.063–0.067)
Bifactor models						
8	Bifactor (TD + PB)	2359.38	254	0.926	0.912	0.060 (0.057–0.062)
9	Model 9 + wording factor (10 items)	1613.07	244	0.952	0.941	0.049 (0.047–0.051)
10	Bifactor (Ext + Int + PB)	2073.80	252	0.936	0.924	0.056 (0.053–0.058)
11	Model 10 + wording factor (10 items)	1156.91	242	0.968	0.960	0.040 (0.038–0.043)

Note. TD = Total Difficulties; PB = Prosocial Behavior; Ext = Externalizing Problems; Int = Internalizing Problems; χ^2 = chi-square; goodness-of-fit statistic; df = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation, with 90% confidence interval (CI); Retained model is marked in bold.
^aWording factor with five positively worded difficulty items.
^bWording factor with all positively worded items.

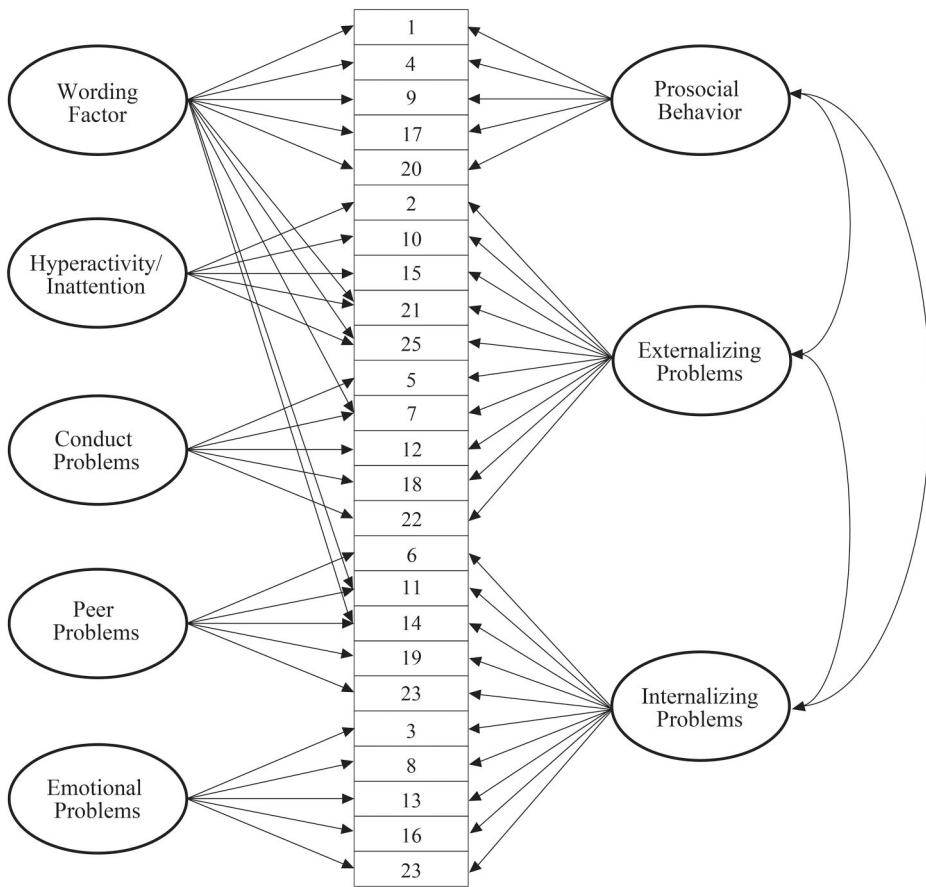


Figure 1. Bifactor model of the SDQ with correlated internalizing and externalizing general factors, domain-specific factors, and a wording-related method factor (Model 11).

Note. Items load on both domain-specific factors and broader Internalizing or Externalizing factors. The Prosocial Behavior factor is modeled separately. The wording factor captures shared variance among the 10 positively worded items.

$H=0.90$) and Internalizing ($\omega_H=0.58$, $H=0.84$) general factors were well-defined and accounted for a substantial proportion of reliable variance ($ECV=0.65$ and 0.55 , respectively). The Externalizing general factor showed strong standardized loadings ($\lambda=0.45\text{--}0.83$, $M=0.65$), indicating that a large share of common variance among the relevant items was attributable to this dimension. The Internalizing general factor likewise demonstrated substantial loadings ($\lambda=0.39\text{--}0.82$, $M=0.54$), supporting its interpretation as a meaningful source of shared variance.

Specific symptom factors

The specific factors showed heterogeneous patterns in their contribution beyond the general dimensions. The Hyperactivity/Inattention factor retained moderate reliable variance

Table 3. Parameter estimates for Model 11 (N = 2332).

Item Content ^a	General symptom factors		Specific symptom factors				Prosocial Behavior	Wording Factor	Residual (1-R ²)
	Externalizing Problems	Internalizing Problems	Hyperactivity/Inattention	Conduct Problems	Emotional Problems	Peer Problems			
2. Restless	0.667		0.650						0.131
10. Fidgety	0.651		0.612						0.202
15. Distracted	0.590		0.613						0.276
21. Thinks ^b	0.453		0.276				0.279		0.640
25. Attention ^b	0.558		0.553				0.222		0.334
5. Loses temper	0.830			0.241					0.252
7. Obedient ^b	0.577			−0.032			0.425		0.485
12. Fights	0.704			0.032					0.503
18. Lies	0.761			0.362					0.289
22. Steals	0.676			0.552					0.238
3. Somatic		0.464			0.298				0.696
8. Worries		0.522			0.652				0.302
13. Unhappy		0.816			0.221				0.285
16. Nervous		0.423			0.386				0.672
24. Fears		0.523			0.651				0.303
6. Solitary		0.385				0.520			0.581
11. One good friend ^b		0.488				0.407	0.423		0.416
14. Liked by others ^b		0.593				0.230	0.526		0.318
19. Bullied		0.659				0.299			0.477
23. Better with adults		0.485				0.497			0.517

(Continued)

Table 3. (Continued).

Item Content ^a	General symptom factors		Specific symptom factors				Wording Factor	Residual (1-R ²)
	Externalizing Problems	Internalizing Problems	Hyperactivity/ Inattention	Conduct Problems	Emotional Problems	Peer Problems		
1. Considerate							-0.475	0.217
4. Shares							-0.394	0.655
9. Helpful							-0.702	0.412
17. Kind to younger							-0.481	0.582
20. Offers to help							-0.556	0.600
Explained common variance (ECV)	0.65	0.55	0.46	0.15	0.41	0.31	0.41	
Scale score reliability (ω)	0.69	0.58	0.35	0.02	0.27	0.21	– ^c	
Construct replicability (H)	0.89	0.84	0.71	0.40	0.64	0.51	– ^c	
Inter-factor correlations			Prosocial behavior		Internalizing problems			
	Externalizing problems		–0.78***		0.74***			
	Internalizing problems		–0.47***		–			

Note. SDQ = Strengths and Difficulties Questionnaire; Significant factor loadings are in bold; *** $p < .001$.

^a Item labels are abbreviations of the original SDQ items.

^b Reverse-scored item.

^c Reliability and construct replicability coefficients are not reported for the wording factor since it was modeled as a nuisance source of covariance rather than as a substantive construct.

and construct replicability ($\omega_{HS} = 0.35$, $H = 0.71$), indicating that some domain-specific variance remained after accounting for the Externalizing factor.

In contrast, the Conduct Problems specific factor was less well defined ($\lambda = -0.032$ – 0.552 , $M = 0.315$), with minimal unique variance ($ECV = 0.15$) and negligible reliable variance beyond the general Externalizing factor ($\omega_{HS} = 0.02$, $H = 0.40$), indicating that variance in this domain was largely subsumed by the broader dimension. Consistent with this pattern, the items that loaded most strongly on the Externalizing factor were Conduct Problems items, including item 5 (Loses temper; $\lambda = 0.83$), item 12 (*Often fights with other children or bullies them*; $\lambda = 0.70$), and item 18 (*Often lies or cheats*; $\lambda = 0.76$).

Within the Internalizing domain, the Emotional Problems specific factor showed moderate loadings ($\lambda = 0.221$ – 0.652 , $M = 0.442$) and modest construct replicability ($\omega_{HS} = 0.27$, $H = 0.64$). Anxiety-related items such as item 8 (*Many worries, often seems worried*; $\lambda = 0.652$) and item 24 (*Many fears, easily scared*; $\lambda = 0.651$) loaded strongly on this factor, whereas item 13 (*Often unhappy, downhearted or tearful*; $\lambda = 0.816$) loaded primarily on the general Internalizing factor.

The Peer Problems factor also demonstrated moderate loadings ($\lambda = 0.230$ – 0.520 , $M = 0.391$) and relatively low construct replicability, suggesting limited reliable variance beyond the general Internalizing factor ($\omega_{HS} = 0.21$, $H = 0.51$). Items such as item 14 (*Generally liked by other children*; $\lambda = 0.230$) and item 19 (*Picked on or bullied by other children*; $\lambda = 0.299$) showed relatively small specific loadings.

Prosocial behavior

The Prosocial Behavior factor showed moderate item-level loadings ($\lambda = 0.301$ – 0.746 , $M = 0.445$), but its interpretation was complicated by substantial method-related variance. Although it accounted for moderate variance ($ECV = 0.44$), scale score reliability and construct replicability was modest ($\omega_{HS} = 0.32$, $H = 0.66$), suggesting that observed subscale scores reflect both substantive item content and systematic response tendencies.

Wording factor

The wording-related method factor captured a substantial proportion of common variance among the positively worded items ($ECV = 0.41$), indicating that item valence contributes meaningfully to their shared covariance. All positively worded items loaded significantly on this factor; prosocial items loaded negatively, whereas positively worded difficulty items loaded positively. The strongest method loadings were observed for item 14 (*Generally liked by other children*; $\lambda = 0.526$), item 7 (*Generally obedient, usually does what adults request*; $\lambda = 0.425$), and item 11 (*Has at least one good friend*; $\lambda = 0.425$). Hyperactivity/Inattention items such as item 21 (*Thinks things through before acting*; $\lambda = 0.279$) and item 25 (*Sees tasks through to the end, good attention span*; $\lambda = 0.222$) showed comparatively smaller wording effects. Taken together, these results indicate that covariance among positively worded items—especially those defining the Prosocial Behavior scale—is partly captured by the wording factor, suggesting that item valence contributes to the observed variance in Prosocial Behavior scores.

Sensitivity analyses

To evaluate whether conclusions regarding the retained bifactor model were sensitive to the assumption that non-target cross-loadings are fixed to zero in CFA, an ESEM model with target rotation was estimated. The model included factors corresponding to Externalizing, Internalizing, and Prosocial Behavior, together with a wording-related method factor (ESEM loading parameters and latent correlations are reported in [Table S3](#)). Model fit was within an acceptable range ($CFI = 0.952$, $TLI = 0.934$, $RMSEA = 0.052$). Allowing cross-loadings did not materially alter the interpretation of the broad Internalizing and Externalizing factors, which remained well defined across specifications, with items showing their largest loadings on the expected domains.

As expected, cross-loadings were observed for some items. For example, item 1 (*Considerate of other people's feelings*) loaded across Externalizing, Prosocial Behavior, and the method factor, and items such as item 12 (*Often fights with other children or bullies them*) and item 11 (*Has at least one good friend*) did not show a clearly dominant loading. In contrast, the Prosocial Behavior factor was not well defined. Prosocial items showed relatively small loadings on their intended factor and tended to load more strongly on the wording-related method factor, while also exhibiting cross-loadings on the Internalizing and Externalizing general factors. Consistent with this, the wording-related method factor remained clearly defined, with its strongest loadings observed for prosocial items.

As an additional sensitivity analysis, a RIIFA model was estimated to examine whether covariance among items could be represented by a general response-style factor. However, model fit was suboptimal ($CFI = 0.939$, $TLI = 0.932$, $RMSEA = 0.077$), and MIs indicated localized areas of misfit, particularly for constraints on the random intercept factor and for additional cross-loadings not represented in the model. For example, several items showed large MIs for the equality-constrained random intercept loadings, including item 7 (*Generally obedient, usually does what adults request*; $MI \approx 306$) and item 1 (*Considerate of other people's feelings*; $MI \approx 129$), as well as for correlations between the random intercept and substantive factors (e.g., RIIFA factor and Internalizing; $MI \approx 136$). These findings suggest that a general response-style factor did not adequately account for the item covariance in the current sample.

Overall, the sensitivity analyses support the robustness of the retained model. Across specifications, the broad Internalizing and Externalizing factors remained well defined, whereas the Prosocial Behavior factor was comparatively weak and closely intertwined with wording-related variance, which was not adequately accounted for by a general response-style factor.

Discussion

The present study examined the latent structure of the SDQ parent-report in a large community sample of Icelandic mothers of children aged 6–12 years using CFA, bifactor, and ESEM models. Across modeling approaches, the most interpretable representation of the SDQ covariance structure was a bifactor model including two broad general factors—Externalizing and Internalizing problems—together with a Prosocial Behavior factor and a wording-related method factor.

The general factors captured a substantial proportion of the common variance and showed satisfactory scale score reliability and construct replicability, suggesting that they are well-defined and likely to generalize across similar samples. At the same time, the specific SDQ factors showed heterogeneous patterns. Conduct Problems and Peer Problems retained limited reliable variance after accounting for the general factors, indicating that these factors contributed relatively little unique information in this community sample. In contrast, Emotional Problems and Hyperactivity/Inattention retained more interpretable levels of specific variance, although this variance remained secondary to the broader dimensions. This pattern is consistent with previous SDQ research demonstrating dominant broad dimensions alongside weaker specific factors (see Caci et al., 2015). The strong association between the broad Externalizing and Internalizing factors further aligns with hierarchical models of psychopathology, in which these domains reflect partially shared underlying liabilities (Kotov et al., 2017; Krueger & Markon, 2006; Lahey et al., 2012, 2017).

The ESEM sensitivity analysis supported these conclusions while clarifying the extent of item-level overlap. The Externalizing and Internalizing factors remained well defined when cross-loadings were estimated, indicating that the main structural conclusions were not dependent on restrictive CFA assumptions. However, several items showed nontrivial cross-loadings, and factor correlations were notably reduced relative to the bifactor-CFA solution. This suggests that part of the strong associations among SDQ domains in more restrictive models reflects item overlap (e.g., Garrido et al., 2020) in addition to the broader latent liabilities.

A different profile emerged for the Prosocial Behavior factor. Although prosocial items formed a distinguishable specific factor in the bifactor model, their standardized loadings on this factor were modest relative to their loadings on the wording-related method factor, scale score reliability was modest and construct replicability was comparatively low. In the ESEM solution, prosocial items showed substantial cross-loadings and did not form a clearly defined substantive factor. Together, these findings indicate that prosocial scores reflect a mixture of substantive content and method-related variance, limiting the interpretability of this factor as a distinct latent construct in the current sample.

In this study, wording-related variance was a prominent feature of the SDQ structure. Explicit modeling of item wording substantially improved model fit, and the wording factor, consisting of the ten positively worded SDQ items, accounted for a considerable proportion of common variance. This pattern is in line with previous studies documenting wording-related variance in the SDQ (e.g., Kankaanpää et al., 2023; McCrory & Layte, 2012; Palmieri & Smith, 2007) and reflects the susceptibility of mixed-valence scales to method effects (e.g., DiStefano & Motl, 2006; Kulas et al., 2019). Importantly, the RIIFA sensitivity analysis did not support a general response-style factor as an adequate explanation for this covariance, suggesting that wording effects are not well captured by a uniform response tendency and instead reflect more structured, item-dependent sources of variance. The method factor may therefore capture a mixture of influences, including item valence and response-related processes (e.g., acquiescent responding or social desirability), although the present analyses do not permit these to be disentangled. Together, these findings indicate that interpretation of the Prosocial Behavior factor in this community sample is complicated by substantial wording-related covariance, underscoring the need to account for this source of variance when working with SDQ data in community populations. Future research is

recommended to clarify the nature of these effects, for example by incorporating external indicators of response tendencies (e.g., social desirability measures).

A further consideration relates to the distributional properties of the SDQ items. In the present study, Conduct Problems items reflecting low base-rate behaviors (e.g., *Lies*, *Fights*, and *Steals*) exhibited pronounced floor effects, whereas positively worded Prosocial Behavior and Peer Problems items (e.g., *Considerate*, *Helpful*, *One good friend*) showed ceiling effects—patterns commonly observed in non-clinical samples. Such restricted variance can attenuate correlations and reduce scale score reliability, and may influence the estimation of latent structure (Flora & Curran, 2004; Greer et al., 2006; Reise et al., 2013). Consistent with this, SDQ subscales—particularly Conduct Problems and Peer Problems—have shown reduced reliability in community samples (see Stone et al., 2010). In this context, the prominence of broad Internalizing and Externalizing factors, alongside weaker specific factors, may partly reflect low base-rate symptom endorsement and increased shared variance across items. Accordingly, distributional properties should be considered when interpreting SDQ factor structures, as they may attenuate associations and reduce reliability.

From an applied perspective, the results support cautious use of SDQ subscale scores in community settings. Broad Internalizing and Externalizing dimensions appear psychometrically more defensible than narrower subscales, particularly Conduct Problems and Peer Problems. While subscale scores may remain descriptively useful, they are best interpreted as indicators of broader liability dimensions and as guides for further assessment rather than as precise measures of distinct constructs. In this context, the Total Difficulties score may provide a useful overall index (Goodman, 2001; Goodman et al., 2010). At the same time, the present findings suggest that two broad dimensions offer a more informative representation of the SDQ in community samples.

Several study limitations should be considered. First, the sample consisted exclusively of mothers from a single Nordic country, with relatively high educational attainment, which may limit generalizability to other respondents (fathers, teachers, youth self-reports) and other cultural contexts. Second, the community-based design entailed low prevalence of several behavioral difficulties and skewed response distributions, which likely influenced factor estimation and specific-factor score reliability. Third, because participants were recruited through schools, responses may be clustered at the classroom or school level. However, as information on clustering was not available, the analyses did not explicitly adjust for complex sampling or intraclass dependence.

Fourth, data were collected via an online survey without formal attention-check items or response-time screening procedures. As with all questionnaire-based data, careless or inconsistent responding cannot be ruled out and may contribute to systematic method-related variance (e.g., Arias et al., 2020). Fifth, precise response rates and information on non-respondents were not available, making it impossible to evaluate systematic differences between respondents and non-respondents.

Finally, although bifactor models can be useful for partitioning variance, they may also overfit when specific factors are weak or sample sizes are large (Bornoalova et al., 2020; Reise et al., 2013), and ESEM solutions remain dependent on rotation method and target specification. The convergence of findings across modeling approaches nonetheless strengthens confidence in the main structural conclusions.

In summary, the present findings suggest that the SDQ parent-report in this community sample primarily reflects two broad general factors—Externalizing and Internalizing problems—while specific factors, particularly Conduct Problems and Peer Problems, show more limited unique variance and reduced scale score reliability after accounting for the general factors. The results also highlight the influence of wording-related covariance and skewed response distributions on the measurement structure of the SDQ. Together, these findings underscore the importance of cautious and context-sensitive interpretation of SDQ subscale scores, particularly in low-risk community settings.

Disclosure statement

The authors declare no conflict of interest.

REFERENCES

- Aichholzer, J. (2014). Random intercept EFA of personality scales. *Journal of Research in Personality*, 53, 1–4. <https://doi.org/10.1016/j.jrp.2014.07.001>
- Arias, V. B., Garrido, L. E., Jenaro, C., Martínez-Molina, A., & Arias, B. (2020). A little garbage in, lots of garbage out: Assessing the impact of careless responding in personality survey data. *Behavior Research Methods*, 52(6), 2489–2505. <https://doi.org/10.3758/s13428-020-01401-8>
- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 397–438. <https://doi.org/10.1080/10705510903008204>
- Bell, B. A., Taylor, S. G., Roberts, A. M., Shi, D., Burgess, K., Hough, C., & Flory, K. (2024). Factor structure of the teacher strengths and difficulties questionnaire in a large community-based sample: An investigation of alternative measurement models. *Assessment*, 31(2), 291–303. <https://doi.org/10.1177/10731911231157627>
- Bonifay, W., Lane, S. P., & Reise, S. P. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science*, 5(1), 184–186. <https://doi.org/10.1177/216770261665706>
- Bornovalova, M. A., Choate, A. M., Fatimah, H., Petersen, K. J., & Wiernik, B. M. (2020). Appropriate use of bifactor analysis in psychopathology research: Appreciating benefits and limitations. *Biological Psychiatry*, 88(1), 18–27. <https://doi.org/10.1016/j.biopsych.2020.01.013>
- Caci, H., Morin, A. J. S., & Tran, A. (2015). Investigation of a bifactor model of the Strengths and Difficulties Questionnaire. *European Child & Adolescent Psychiatry*, 24(10), 1291–1301. <https://doi.org/10.1007/s00787-015-0679-3>
- DiStefano, C., & Morgan, G. B. (2014). A comparison of diagonal weighted least squares robust estimation techniques for ordinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(3), 425–438. <https://doi.org/10.1080/10705511.2014.915373>
- DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling: A Multidisciplinary Journal*, 13(3), 440–464. https://doi.org/10.1207/s15328007sem1303_6
- Dueber, D. M. (2022). *BifactorIndicesCalculator: Compute bifactor indices from bifactor CFA output*. R package version 0.2.2. <https://CRAN.R-project.org/package=BifactorIndicesCalculator>
- Duinhof, E. L., Lek, K. M., de Looze, M. E., Cosma, A., Mazur, J., Gobina, I., Wüstner, A., Vollebergh, W. A. M., & Stevens, G. W. J. M. (2019). Revising the self-report strengths and difficulties questionnaire for cross-country comparisons of adolescent mental health problems: The SDQ-R. *Epidemiology and Psychiatric Sciences*, 29, e35. <https://doi.org/10.1017/S2045796019000246>
- Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, 9(4), 466–491. <https://doi.org/10.1037/1082-989X.9.4.466>
- Garrido, L. E., Barrada, J. R., Aguasvivas, J. A., Martínez-Molina, A., Arias, V. B., Golino, H. F., Legaz, E., Ferris, G., & Rojo-Moreno, L. (2020). Is small still beautiful for the strengths and difficulties

- questionnaire? Novel findings using exploratory structural equation modeling. *Assessment*, 27(6), 1349–1367. <https://doi.org/10.1177/1073191118780461>
- Goodman, A., Lamping, D. L., & Ploubidis, G. B. (2010). When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): Data from British parents, teachers and children. *Journal of Abnormal Child Psychology*, 38(8), 1179–1191. <https://doi.org/10.1007/s10802-010-9434-x>
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 38(5), 581–586. <https://doi.org/10.1111/j.1469-7610.1997.tb01545.x>
- 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. <https://doi.org/10.1097/00004583-200111000-00015>
- Greene, A. L., Eaton, N. R., Li, K., Forbes, M. K., Krueger, R. F., Markon, K. E., Waldman, I. D., Cicero, D. C., Conway, C. C., Docherty, A. R., Fried, E. I., Ivanova, M. Y., Jonas, K. G., Latzman, R. D., Patrick, C. J., Reininghaus, U., Tackett, J. L., Wright, A. G. C., & Kotov, R. (2019). Are fit indices used to test psychopathology structure biased? A simulation study. *Journal of Abnormal Psychology*, 128(7), 740–764. <https://doi.org/10.1037/abn0000434>
- Greer, T., Dunlap, W. P., Hunter, S. T., & Berman, M. E. (2006). Skew and internal consistency. *Journal of Applied Psychology*, 91(6), 1351–1358. <https://doi.org/10.1037/0021-9010.91.6.1351>
- Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In *Structural equation modeling: Present and future* (pp. 195–216). Scientific Software International.
- Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika*, 2(1), 41–54. <https://doi.org/10.1007/BF02287965>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Kankaanpää, R., Töttö, P., Punamäki, R. L., & Peltonen, K. (2023). Is it time to revise the SDQ? The psychometric evaluation of the Strengths and Difficulties Questionnaire. *Psychological Assessment*, 35(12), 1069–1084. <https://doi.org/10.1037/pas0001265>
- Kóbor, A., Takács, Á., & Urbán, R. (2013). The bifactor model of the Strengths and Difficulties Questionnaire. *European Journal of Psychological Assessment*, 29(4), 299–307. <https://doi.org/10.1027/1015-5759/a000160>
- Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., ... Zimmerman, M. (2017). The hierarchical taxonomy of psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, 126(4), 454–477. <https://doi.org/10.1037/abn0000258>
- Krueger, R. F., & Markon, K. E. (2006). Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. *Annual Review of Clinical Psychology*, 2(1), 111–133. <https://doi.org/10.1146/annurev.clinpsy.2.022305.095213>
- Kulas, J. T., Klahr, R., & Knights, L. (2019). Confound it!: Social desirability and the “reverse-scoring” method effect. *European Journal of Psychological Assessment*, 35(6), 855–867. <https://doi.org/10.1027/1015-5759/a000459>
- Lahey, B. B., Applegate, B., Hakes, J. K., Zald, D. H., Hariri, A. R., & Rathouz, P. J. (2012). Is there a general factor of prevalent psychopathology during adulthood? *Journal of Abnormal Psychology*, 121(4), 971–977. <https://doi.org/10.1037/a0028355>
- Lahey, B. B., Krueger, R. F., Rathouz, P. J., Waldman, I. D., & Zald, D. H. (2017). A hierarchical causal taxonomy of psychopathology across the life span. *Psychological Bulletin*, 143(2), 142–186. <https://doi.org/10.1037/bul0000069>
- Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, 10(1), 85–110. <https://doi.org/10.1146/annurev-clinpsy-032813-153700>
- Maydeu-Olivares, A., & Coffman, D. L. (2006). Random intercept item factor analysis. *Psychological Methods*, 11(4), 344–362. <https://doi.org/10.1037/1082-989X.11.4.344>

- McCrary, C., & Layte, R. (2012). Testing competing models of the Strengths and Difficulties Questionnaire's (SDQ's) factor structure for the parent-informant instrument. *Personality and Individual Differences*, 52(8), 882–887. <https://doi.org/10.1016/j.paid.2012.02.011>
- Ortuño-Sierra, J., Fonseca-Pedrero, E., Aritio-Solana, R., Velasco, A. M., de Luis, E. C., Schumann, G., Cattrell, A., Flor, H., Nees, F., Banaschewski, T., Bokde, A., Whelan, R., Buechel, C., Bromberg, U., Conrod, P., Frouin, V., Papadopoulos, D., Gallinat, J., Garavan, H., ... Lawrence, C. IMAGEN Consortium. (2015). New evidence of factor structure and measurement invariance of the SDQ across five European nations. *European Child & Adolescent Psychiatry*, 24(12), 1523–1534. <https://doi.org/10.1007/s00787-015-0729-x>
- Palmieri, P. A., & Smith, G. C. (2007). Examining the structural validity of the Strengths and Difficulties Questionnaire (SDQ) in a U.S. sample of custodial grandmothers. *Psychological Assessment*, 19(2), 189–198. <https://doi.org/10.1037/1040-3590.19.2.189>
- R Core Team. (2022). *R: A language and environment for statistical computing (Version 4.2)*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47(5), 667–696. <https://doi.org/10.1080/00273171.2012.715555>
- Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, 95(2), 129–140. <https://doi.org/10.1080/00223891.2012.725437>
- Revelle, W. (2022). *psych: Procedures for psychological, psychometric, and personality research (Version 2.2.3)*. Northwestern University. <https://CRAN.R-project.org/package=psych>
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, 21(2), 137–150. <https://doi.org/10.1037/met0000045>
- Rosseel, Y., Jorgensen, T., De Wilde, L. (2023). *lavaan: Latent Variable Analysis*. R package version 0.6-15. <https://CRAN.R-project.org/package=lavaan>
- Skoczeń, I., Rogoza, R., Maćkiewicz, M., Najderska, M., & Ciecuch, J. (2018). Investigating the structural model of the Strengths and Difficulties Questionnaire. *European Journal of Psychological Assessment*, 34(5), 312–320. <https://doi.org/10.1027/1015-5759/a000344>
- Stone, L. L., Otten, R., Engels, R. C., Vermulst, A. A., & Janssens, J. M. (2010). Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4-to 12-year-olds: A review. *Clinical Child and Family Psychology Review*, 13(3), 254–274. <https://doi.org/10.1007/s10567-010-0071-2>
- Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.) Pearson.
- Thorlacius, Ö., & Gudmundsson, E. (2015). Assessment of children's emotional adjustment: Construction and validation of a new instrument. *Child: Care, Health and Development*, 41(5), 762–771. <https://doi.org/10.1111/cch.12214>
- Thorlacius, Ö., & Gudmundsson, E. (2019). The effectiveness of the children's emotional adjustment scale (CEAS) in screening for mental health problems in middle childhood. *School Mental Health*, 11(3), 400–412. <https://doi.org/10.1007/s12310-018-9296-x>
- Valborgarson, A. (2016). Assessment of children's behavior and emotional well-being [unpublished data].
- Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1–67. <https://doi.org/10.18637/jss.v045.i03>
- van de Looij-Jansen, P. M., Goedhart, A. W., de Wilde, E. J., & Treffers, P. D. (2011). Confirmatory factor analysis and factorial invariance analysis of the adolescent self-report Strengths and Difficulties Questionnaire: How important are method effects and minor factors? *British Journal of Clinical Psychology*, 50(2), 127–144. <https://doi.org/10.1348/014466510X498174>
- 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. <https://doi.org/10.1027/1015-5759.22.3.189>
- 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, and Allied Disciplines*, 49(12), 1304–1312. <https://doi.org/10.1111/j.1469-7610.2008.01942.x>

Vugteveen, J., de Bildt, A., Theunissen, M., Reijneveld, S. A., & Timmerman, M. (2021). Validity aspects of the strengths and difficulties questionnaire (SDQ) adolescent self-report and parent-report versions among Dutch adolescents. *Assessment*, 28(2), 601–616. <https://doi.org/10.1177/1073191119858416>