Scandinavian Journal of Psychology, 2022 DOI: 10.1111/sjop.12815

Empirical Article

Internalizing problems, externalizing problems, and prosocial behavior - three dimensions of the strengths and difficulties questionnaire (SDQ): A study among South African adolescents

LEIF EDVARD AARØ,[1] EUGENE L. DAVIDS,[2,3] CATHERINE MATHEWS,[2,4] ANNEGREET GERA WUBS,[5] OTTO R.F. SMITH[1] and PETRUS J. DE VRIES[2,3]

[1]*Department of Health Promotion, Norwegian Institute of Public Health, Bergen, Norway*
[2]*Adolescent Health Research Unit, Department of Psychiatry and Mental Health, University of Cape Town, Cape Town, South Africa*
[3]*Division of Child & Adolescent Psychiatry, University of Cape Town, Cape Town, South Africa*
[4]*Health Systems Research Unit, South African Medical Research Council, Cape Town, South Africa*
[5]*Department of Health Promotion and Development, University of Bergen, Bergen, Norway*

Aarø, L. E., Davids, E. L., Mathews, C., Wubs, A. G., Smith, O. R. F. & de Vries P. J. (2022). Internalizing problems, externalizing problems, and prosocial behavior - three dimensions of the strengths and difficulties questionnaire (SDQ): A study among South African adolescents. *Scandinavian Journal of Psychology.*

This study aimed to examine the factor structure of the Strengths and Difficulties Questionnaire Self-Report version (SDQ-S), its psychometric properties and measurement invariance by gender and language spoken at home, among secondary school students in Western Cape, South Africa. A sample of 3,542 adolescents in Grade 8 (*Mean age* = 13.7 years) completed the SDQ-S in a three-language questionnaire (Afrikaans, English and isiXhosa). The data were collected from 42 secondary schools in Cape Town, South Africa. Confirmatory factor analyses with the WLSMV estimator with adjustment for cluster effects (schools) were applied. The SDQ-S was originally developed to cover five domains: four "difficulty" domains (hyperactivity/inattention, emotional, conduct, and peer) and one "strengths" domain (prosocial behavior). When the five factors were tested on the data for the current study, poor fit was obtained. After excluding four items, a three-factor solution with no cross-loadings and no correlated error terms obtained acceptable fit. The results are consistent with previous studies. Strong measurement invariance across genders and language spoken at home was confirmed. In studies of community samples, the use of the SDQ-S scale as an instrument with a three-factor dimension (internalizing, externalizing and prosocial) may be more appropriate than using the original five-factor model.

Key words: Adolescents, dimensions, measurement invariance, South Africa, strengths and difficulties questionnaire (SDQ).

Leif Edvard Aarø, Norwegian Institute of Public Health, Department of Health Promotion, Zander Kaaesgt. 7, NO-5015 Bergen, Norway; Tel: +47 – 53 20 41 02; phone: +47 – 45 88 98 99; e-mail: leif.edvard.aaro@fhi.no

INTRODUCTION

Approximately 85% of the global population live in low- and middle-income countries (LMICs) which is home to just over 80% of people with mental health and substance use disorders (Rathod, Pinninti, Irfan *et al*., 2017). Added to this mental health burden, close to 75% of these diagnoses have their onset before age 24 (Kessler, Amminger, Aguilar-Gaxiola, Alonso, Lee & Ustun, 2007). This age group (24 years and younger) also accounts for close to 40% of the global population of which majority are from low- and middle-income countries (Patel, Flisher, Nikapota & Malhotra, 2008), contributing more than a quarter of the disability-adjusted life years for mental disorders among this age group (Lu, Li & Patel, 2018). Children and adolescents from low- and middle income countries often have limited to no access to mental health care services (Lu *et al*., 2018) further adding pressure to the already resource stricken economies in these countries.

Child and adolescent mental health (CAMH) in these LMICs are synonymous with human resources that have limited skilled professionals, high workloads and service demands, socio-economic and health system factors, linguistic and culturally inappropriate measures and interventions (Klasen & Crombag, 2013; Patel *et al*., 2008). In South Africa, as an example of a LMIC, children and adolescents are at significant risk for mental health problems given the high rates of poverty, violence and the growing HIV epidemic, which persists in the country (Mellins, Xu, Nestadt *et al*., 2018). Similar high risks exist in other low- and middle-income countries (Lu, Black & Richter, 2016). Child and adolescent mental health (CAMH) problems as well as the ever-growing treatment gap which exist between the need for and access to mental health services in the country is often a result of limited resources and availability of qualified staff. Furthermore, an additional barrier to CAMH service provision is the limited availability of translated assessment measures that are reliable, valid and easy to use among non-mental health professionals (de Vries, Davids, Mathews & Aarø, 2018; Hoosen, Davids, de Vries & Shung-King, 2018; Lund, Kleintjies, Kakuma, Flisher & MHaPP Research Programme Consortium, 2010; Mellins *et al*., 2018). The Strengths and Difficulties Questionnaire (SDQ) is a common screening tool used among lay professionals that is brief and easy to administer in the context of CAMH (Garrido, Barrada, Agusvivas *et al*., 2018; Goodman, Lamping & Ploubidis, 2010).

The SDQ is a widely used behavioral screening tool which allows for a quick and easy to use measurement of mental health difficulties and competencies (Hoosen *et al*., 2018). It was

Section Editor: Dr Ewa Mörtberg

originally developed for the measurement of five aspects related to mental health screening namely four "difficulty" domains (hyperactivity/inattention, emotional, conduct, and peer) and one "strengths" domain (prosocial behavior) (Goodman, 1997).

In his 1997 article on SDQ, Goodman carried out analyses of the dimensionality of all three versions of the scale. He used data from a sample of 5–15 year olds, their parents, and teachers in the UK. The youth version (SDQ-S) was completed by the 11–15 year-olds only. The analyses suggested that the five-factor structure was generally confirmed, although some cross-loadings were significant, particularly for the youth version. The Cronbach alpha scores ranged from 0.70 to 0.88 for the teacher version, 0.57 to 0.85 for the parent version and from 0.41 (peer problems subscale) to 0.80 for the youth version (Goodman, 1997).

Subsequent analyses of the dimensionality of the SDQ in different countries have been carried out. The findings of the dimensionality suggest that the five-factor structure has received some support from studies in Germany (Becker, Hagenberg, Roessner, Woerner & Rothenberger, 2004) and Norway (Van Roy, Veenstra & Clench-Aas, 2008). Van de Looij-Jansen, Goedhart, de Wilde and Treffers (2011) obtained good fit only after allowing cross-loadings and correlated error terms and found some support for a four-factor model. Under the five-factors model, an extended Prosocial-factor (under the term "positive construal factor") included a number of cross-loadings to the other factors. Percy and colleagues have pointed out several deviations from the hypothesized five-factor structure including a lack of unidimensionality within some subscales, cross-loadings, and items failing to load onto any factor (Percy, McCrystal & Higgins, 2008). Koskelainen, Sourande, and Vauras (2001) probably were the first ones to suggest that the SDQ should be used as a three-dimensional instrument. They, however, did not present the relevant tables and statistics in support of their suggestion. Factor analyses of SDQ data collected among parents of children aged 4 to 17 years in the United States confirmed a three-factor structure (Dickey & Blumberg, 2004). The terms used to describe the three-factor structure were "externalization problems," "internalization problems," and a third, which was termed a "positive construal factor" (Dickey & Blumberg, 2004). Further support of the three-factor structure of the scale has been provided in studies based on the SDQ-S (child and adolescent self-reports) from the United States (Ruchkin, Jones, Vermeiren & Schwab-Stone, 2008), from Italy (Di Riso, Salcuni, Chessa, Raudino, Lis. & Altoe, 2010), and from Spain (Ortuño-Sierra, Fonseca-Pedrero, Paino, Sastre i Riba & Muñiz, 2015). The three-factor structure has also received support in a study based on teachers' reports from Saudi-Arabia and Oman (El-Keshky & Emam, 2015) and in a study based on parents' reports from Spain (Gómez-Beneyto, Nolasco, Moncho *et al*., 2013). In South Africa, like many LMICs the SDQ is used for screening of mental health challenges to assist in the CAMH treatment gap which exists. The SDQ-S, however, was found to have poor support for the five-factor structure in South Africa, and only after some modifications was an acceptable fit found (de Vries *et al*., 2018). These earlier findings led to the analyses and the new three-factor model presented in the present paper.

Further analyses of the psychometric properties of the original five dimensions of SDQ have raised some concern about the quality of the scale, particularly for cross-cultural comparison. Rønning, Handegaard, Sourander and Mørch (2004) identified several items with high loadings on more than one factor during their analyses of the psychometric properties. Hagquist (2007) also found problems in all five dimensions of the SDQ. The conduct problems subscale proved to be particularly problematic. Some items may have rather different meanings in different contexts, and there are items consisting of different statements with divergent meanings. Hagquist suggested that the problem of ambiguity of some items could be reduced by splitting each item that contained more than one single meaning into separate items. Others, on the other hand, have cautioned against the use of the SDQ completely, suggesting that the SDQ has a "weak and unstable factorial structure," lacks robustness in its factorial structure, and have findings supporting either a five-factor structure with a relatively weak fifth factor or a four-factor structure (Garrido *et al*., 2018).

In LMICs like South Africa, the risk for child and adolescent mental health problems is considered to be high, and there is a serious CAMH treatment gap. Under such circumstances, understanding the use of the SDQ-S and its psychometric properties, measurement invariance as well as factor structure becomes particularly important. In a scoping review of the use of the SDQ in Africa (Hoosen *et al*., 2018) it was found that the SDQ was frequently used, but inappropriately – with some using certain subscales of the SDQ and others allowing for the use outside of the recommended age guidelines by the developers. Further, the review highlighted that little was known about the psychometric properties of the SDQ in African contexts, recommending a need for an evaluation of the SDQ in the various African languages (as in other diverse settings), internal consistency, factor structure and local cut-off scores allowing for cultural equivalences of the instrument.

There is a dramatic scarcity of psychiatrists and clinical psychologists in South Africa, and even more so in other parts of the continent. In South Africa SDQ is still widely used in health care settings because it is a freely-available tool. It is for instance used to make decisions in "stepped care" contexts. In such a clinical care pathway the SDQ can be used safely and appropriately mindful of the limitations in terms of its psychometric properties. It has also been recommended by the South African Integrated School Health Programme (ISHP) as a screening tool for young people. This is more of a problem, since its dimensionality is unclear, and since norms defining high levels of problems established in high income countries may be quite misleading.

In a psychometric evaluation of the SDQ in South African adolescents, we previously highlighted that the five-factor structure did not seem very well supported by our data and suggested further exploration of the factor structure of the SDQ-S (de Vries*et al*., 2018). These earlier findings led to the analyses and the new three-factor model presented in the present paper.

Many adult mental health problems have been established during childhood and adolescence highlighting the need for early screening and detection to promote early intervention and treatment. With current contradictions about the use of the SDQ and its factor structure the purpose of the current study is to examine the factor structure of the SDQ-S scale, its psychometric properties and measurement invariance by gender and language spoken at home using data collected among secondary school students in Western Cape, South Africa.

METHOD

Participants

The data used in this study forms part of baseline data collected as part of a larger cluster, randomized controlled trial, PREPARE (Promoting Sexual and Reproductive Health among Adolescents in Southern and Eastern Africa) (Aarø, Mathews, Kaaya *et al*., 2014). We randomly sampled 42 public high schools from the database of high schools in the Western Cape Province in South Africa. One school dropped out of the study at a later stage and was therefore not included in the analyses presented here. Participants were adolescents in Grade 8 (mean age 13.7 years) attending the sampled schools. For more details of the PREPARE trial methods, see Aarø *et al* (2014) and Mathews, Eggers, Townsend *et al*. (2016). We invited 6,244 students to participate in the PREPARE trial and 3,451 (55.3%) returned signed parental/caregiver consent forms, gave assent and participated in the baseline survey in February and March 2013. The non-responders included 69 students and 281 parents who declined permission for their child to participate and the remainder were students who did not return signed parental consent forms.

Measures

The 25-item SDQ instrument requires the tool administrator to respond to items as being either "not true', "somewhat true," and "certainly true," with scores of 0, 1 and 2 for unfavorably phrased problem items and 2, 1 and 0 to prosocial items and favorably phrased problem items. A higher score obtained for the hyperactivity/inattention, emotional symptoms, conduct and peer problems subscales suggests more significant problems, while a high prosocial score suggest improved or better prosocial behavior. The scale is available in a self-report version to be completed by adolescents themselves. There is also a parent/caregiver version and a teacher version.

In the present paper we use data from the baseline data collection of the PREPARE study. All questionnaires were provided in the three languages commonly spoken in the Western Cape Province (Afrikaans, English and isiXhosa), and were printed in an adolescent-friendly format resembling a "teen magazine." The self-report version of the Strengths and Difficulties Questionnaire (SDQ-S) was included as part of the study questionnaire at baseline and 12 months. The analyses described here were carried out on data from the baseline survey only.

We used the English version of the SDQ-S and undertook standard procedures for translation and back-translation to develop the Afrikaans and isiXhosa versions, as required by the authors and developers of the tool. The translation phase also included an expert panel review of the Afrikaans and isiXhosa versions for cultural and pragmatic appropriateness, and to assess whether the translated words and ideas accurately reflected the original version of the SDQ-S. When uncertainties arose, we contacted the original tool developer of the SDQ for clarification and guidance.

Procedure

The baseline survey was conducted in classrooms during school hours, and completion of the questionnaire took on average 45 min. The SDQ scale was only one among a large number of questions and scales in the questionnaire used in the PREPARE study. Data collection was carried out by a team of research assistants consistent with a procedure that ensured confidentiality to all participants. Teachers were not present during data collection.

Data analysis

The descriptive analyses, which included an analysis of the Cronbach alpha values, were conducted using IBM SPSS Statistics 22. All analyses of dimensionality of the SDQ-S scale, which included the confirmatory factor analyses with no restrictions on inter-factor correlations and control for cluster effects, were analysed using Mplus Version 7. Mplus Version 7 was also used for testing the measurement invariance by gender and by language spoken at home. All indicators were defined as (ordered) categorical and the estimator used was weighted least squares mean and variance adjusted (WLSMV). Model fit was assessed with Chi Square (χ^2), root mean square error of approximation (RMSEA), comparative fit index (CFI), and Tucker–Lewis index (TLI). When models were adapted to the data by allowing cross-loadings or correlated error terms, four criteria were used: (1) modification chi square (χ^2) values standing out as higher than most other modification chi square (χ^2) values (figure-ground contrast); (2) correlated error terms equal to or higher than 0.25; (3) cross-loadings higher than 0.25; and (4) interpretability.

RESULTS

We sampled 6,244 adolescents in 41 schools, and 3,451 (55.3%) obtained signed parental consent and assented to participate. The non-responders included 69 students and 281 parents who declined, and the remaining were students who did not bring back signed parental consent forms (Mathews *et al*., 2016).

One-way frequency distributions for all SDQ-S items are shown in Table 1. Among the 3,542 participants, 102 (2.9%) did not answer any of the 25 SDQ-S items. Ninety two percent had valid answers on at least 23 out of the 25 SDQ-S items. Item 10 ("I am constantly fidgeting or squirming'" had a particularly high proportion of missing responses. One possible explanation could have been students' lack of familiarity with the words "fidgeting" and "squirming." Only one item was strongly skewed: item 11 ("I fight a lot. I can make other people do what I want.") (skewness: −2.844, kurtosis: 7.330). These two items deserve special attention when further presenting and discussing the results of this study.

When testing the five-factor model of the SDQ-S scale (emotional symptoms, conduct problems, hyperactivity / inattention, peer problems, and prosocial behavior), close to adequate fit was obtained only for RMSEA (0.052). Poor fit was obtained when using other standard fit indices (CFI = 0.629; TLI = 0.580). For two of the subscales (emotional symptoms; prosocial behavior) all factor loadings were higher than 0.40, while, the other three subscales only had between two or three loadings that were higher than 0.40. Alpha values were 0.69 (prosocial behavior), 0.59 (emotional symptoms), 0.51 (hyperactivity / inattention), 0.37 (conduct problems), and 0.29 (peer problems).

After adding seven cross-loadings (loading on "wrong" factors) (Fig. 1), based on modification indices, better fit was obtained (RMSEA = 0.021; CFI = 0.942; TLI = 0.933). Loadings higher than 0.40 were obtained for four out of the seven cross-loadings. Prosocial Behavior "borrowed" two items from each of the three factors: Conduct Problems, Hyperactivity/inattention, and Peer Problems. Emotional Symptoms "borrowed" one item from the Conduct Problems subscale.

A three-factor model (Externalizing, Internalizing, Prosocial) based on 21 out of the original 25 SDQ-S items, with seven items in each, obtained acceptable fit (Chi square = 695.042; d.f. = 186; $p < 0.001$) (RMSEA = 0.028; CFI = 0.911; TLI = 0.900) (Fig. 2). Two items were excluded since they did not fit well conceptually into any of the three factors, namely: Item 21 ("Think things out before acting") and item 25 ("Sees

Table 1. *Frequency distributions of all single SDQ items*

	Not true %	Somewhat true %	Certainly true %	Total %	Total N
1 – I try to be nice to other people. I care about their feelings.	5.8	30.5	63.7	100.0	3,370
2 – I am restless; I cannot stay still for long.	48.8	30.4	20.9	100.0	3,324
3 – I get a lot of headache, stomach aches or sickness.	48.0	31.1	20.9	100.0	3,346
4 – I usually share with others (food, games, pens etc.).	9.0	25.4	65.6	100.0	3,353
5 – I get very angry and often lose my temper.	46.8	31.0	22.2	100.0	3,343
6 – I am usually on my own. I generally play alone or keep to myself.	69.8	16.9	13.3	100.0	3,349
7 – I usually do as I am told.	9.8	38.6	51.6	100.0	3,371
8 – I worry a lot.	35.9	34.2	29.9	100.0	3,340
9 – I am helpful if someone is hurt, upset of feeling ill.	9.1	29.4	61.5	100.0	3,381
10 – I am constantly fidgeting or squirming.	66.0	27.1	6.9	100.0	3,185
11 – I have one good friend or more.	14.6	16.2	69.2	100.0	3,362
12 – I fight a lot. I can make other people do what I want.	87.2	9.1	3.7	100.0	3,317
13 – I am often unhappy, down-hearted or tearful.	54.5	29.4	16.2	100.0	3,341
14 – Other people my age generally like me.	10.9	24.6	54.5	100.0	3,339
15 – I am easily distracted, I find it difficult to concentrate.	50.5	32.9	16.6	100.0	3,315
16 – I am nervous in new situations. I easily lose confidence.	42.6	33.9	23.6	100.0	3,337
17 – I am kind to younger children.	7.1	23.4	69.6	100.0	3,359
18 – I am often accused of lying or cheating.	53.4	29.1	17.5	100.0	3,312
19 – Other children or young people pick on me or bully me	66.2	20.5	13.2	100.0	3,336
20 – I often volunteer to help others (parents, teachers, children).	10.4	36.6	53.0	100.0	3,355
21 – I think before I do things	7.7	28.5	63.7	100.0	3,334
22 – I take things that are not mine from home, school or elsewhere.	79.2	14.2	6.6	100.0	3,327
23 – I get on better with adults than with people my age.	42.4	31.0	26.6	100.0	3,329
24 – I have many fears, I am easily scared	41.6	34.0	24.5	100.0	3,339
25 – I finish the work I'm doing. My attention is good.	7.5	33.7	58.8	100.0	3,369

tasks through to the end, good attention span"). Two more items were also excluded because they obtained low loadings and contributed negatively to Cronbach's alpha: Item 11 ("Has at least one good friend") and item 23 ("Gets on better with adults than with other children"). Alpha values for the three scales were 0.57 (Externalizing), 0.63 (Internalizing) and 0.72 (Prosocial).

Based on modification indices, model fit could be improved by allowing two cross-loadings on the Internalizing factor: "Fight, dominate others" (coefficient = −0.36) and "Angry, lose temper" (coefficient = 0.28) and one cross-loading on the Externalizing factor "I worry a lot" (coefficient = −0.32). Model fit was further improved by including one correlation between error terms: "Restless" with "Fidgeting or squirming" (standardized coefficient 0.27) (Fig. 3). After these modifications, fit indices were: RMSEA = 0.025; CFI = 0.933; TLI = 0.922.

The "Prosocial" factor had the highest alpha value (as already mentioned above) (0.72). All standardized factor loadings were higher than 0.40 (actual range 0.49 to 0.71 for the plain as well as for the adjusted model). There were no cross-loadings involving the prosocial factor.

The internalizing subscale had the second highest alpha value of 0.63 and all loadings except one (0.39) were 0.40 or higher (actual range 0.39–0.62 for the plain model and 0.38–0.90 for the adjusted model). One of the items reflected the externalizing factor. The externalizing subscale had an alpha value of 0.57, loadings in the plain three factor model ranging from 0.40 to 0.57, and loadings in the modified model ranging from 0.23 to 0.77. Two items also loaded onto the Internalizing factor and there was one correlation higher than 0.25 between error terms.

Two of the factors, Internalizing and Externalizing, were highly correlated. In the plain three-factor model, the correlation between the two factors was 0.73. In the modified model, it was even higher at 0.80. High intercorrelations among factors were also observed in the modified five-factor model: 0.81 (peer problems and emotional symptoms), 0.73 (hyperactivity / inattention and conduct problems), and 0.67 (hyperactivity / inattention and emotional symptoms).

In order to examine measurement invariance by gender, as a first step, models were developed for boys and girls separately. Interestingly, the clean three-factor models were quite similar in terms of loadings, associations among factors and fit indices (Boys: $\chi^2 = 382.272$, d.f. = 186; RMSEA = 0.027; CFI = 0.924; TLI = 0.914. Girls: RMSEA = 0.025; CFI = 0.932; TLI = 0.922). In a multi-group analysis with gender as the group variable, a model with fully restricted unstandardized loadings and inter-factor associations, acceptable fit was obtained ($\chi^2 = 868.938$, d.f. = 408; RMSEA = 0.026; CFI = 0.921; TLI = 0.919).

Measurement invariance was tested by gender for the clean, unmodified three-factor model (Table 2). With the sample size at hand in this study, all differences between models tested (configural, metric and scalar) were, unsurprisingly, significant. There was, however, no difference between the models with regard to RMSEA, and only marginal differences on CFI and TLI. The TLI statistics actually improved marginally with increasing model restrictiveness (0.912, 0.913, and 0.914 for the configural, metric and scalar models respectively). There was, in other words, a high level of measurement invariance.

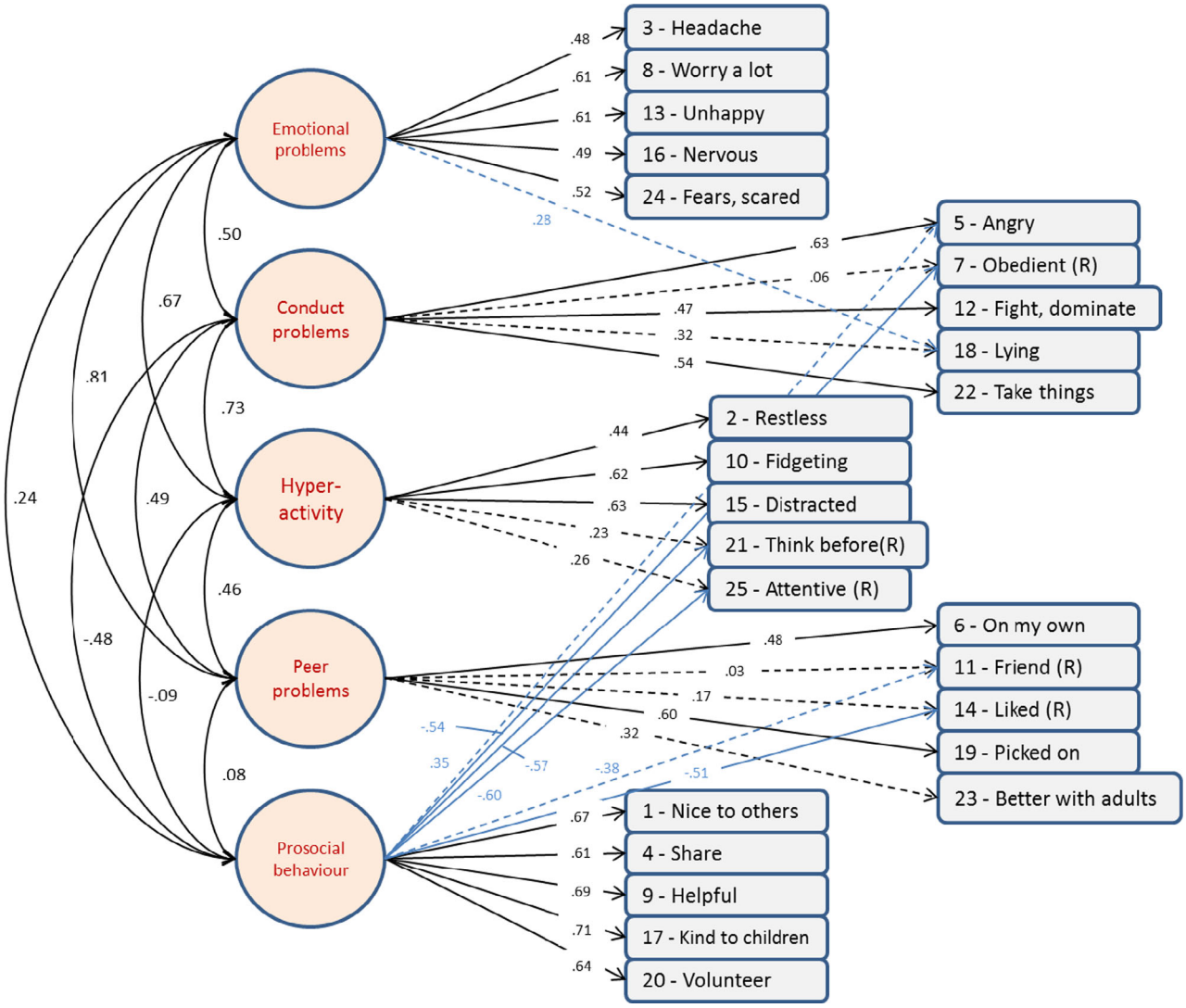

Fig. 1. Five factors after inclusion of significant cross-loadings. WLSMV estimator with Mplus version 7. YX standardized coefficients. N = 3,440; Clusters = 41; Chi square = 637.540, d.f. = 258, p = .0000; RMSEA = .021; CFI = 0.942; TLI = 0.933. (R) means coding reversed. Fig 1 is also presented in de Vries *et al*., 2018.

Measurement invariance was also tested for the modified 3-factor model (Table 3). Measurement invariance testing of the metric model specifically is not possible when the model is based on polytomous categorical variables when there are variables loading on more than one factor. Scalar versus configural invariance could, however, be tested. There was no change in RMSEA (0.024) and only marginal change in CFI (from 0.938 to 0.932) and TLI (from 0.932 to 0.929). This confirmed strong invariance.

The same procedure was applied for testing of invariance across languages (isiXhosa, English and Afrikaans). The clean three-factor models were similar (isiXhosa: χ^2 = 279.152, d.f. = 186; RMSEA = 0.022; CFI = 0.930; TLI = 0.921; English: χ2 = 295.080, d.f. = 186; RMSEA = 0.030; CFI = 0.905; TLI = 0.892. Afrikaans: χ2 = 346.617, d.f. = 186; RMSEA = 0.025; CFI = 0.921; TLI = 0.921). In a multi-group analysis with language as the group variable, a model with fully restricted unstandardized loadings and inter-factor associations, with acceptable fit was obtained: (χ2 = 1059.124, d.f. = 633; RMSEA = 0.026; CFI = 0.908; TLI = 0.908).

Measurement invariance was tested by language for the clean, unmodified three-factor model (Table 4). There was no difference between the models with regard to RMSEA, and only marginal differences on CFI and TLI. There was, in other words, a high level of measurement invariance. Measurement invariance was also tested for the modified 3-factor model (Table 5). There were only negligible changes in fit indices.

Finally, associations between the three SDQ-S subscales and a number of sociodemographic factors were estimated for simple SDQ-S scores, latent variables from the plain three factor solution and latent variables based on the modified three factor solution (Table 6). All significant associations were in the same direction and approximately of the same size. Generally, the latent variables obtained stronger associations than scores. This is due to properties of the WLSMV estimator as well as the correction for attenuation built into analyses with latent variables.

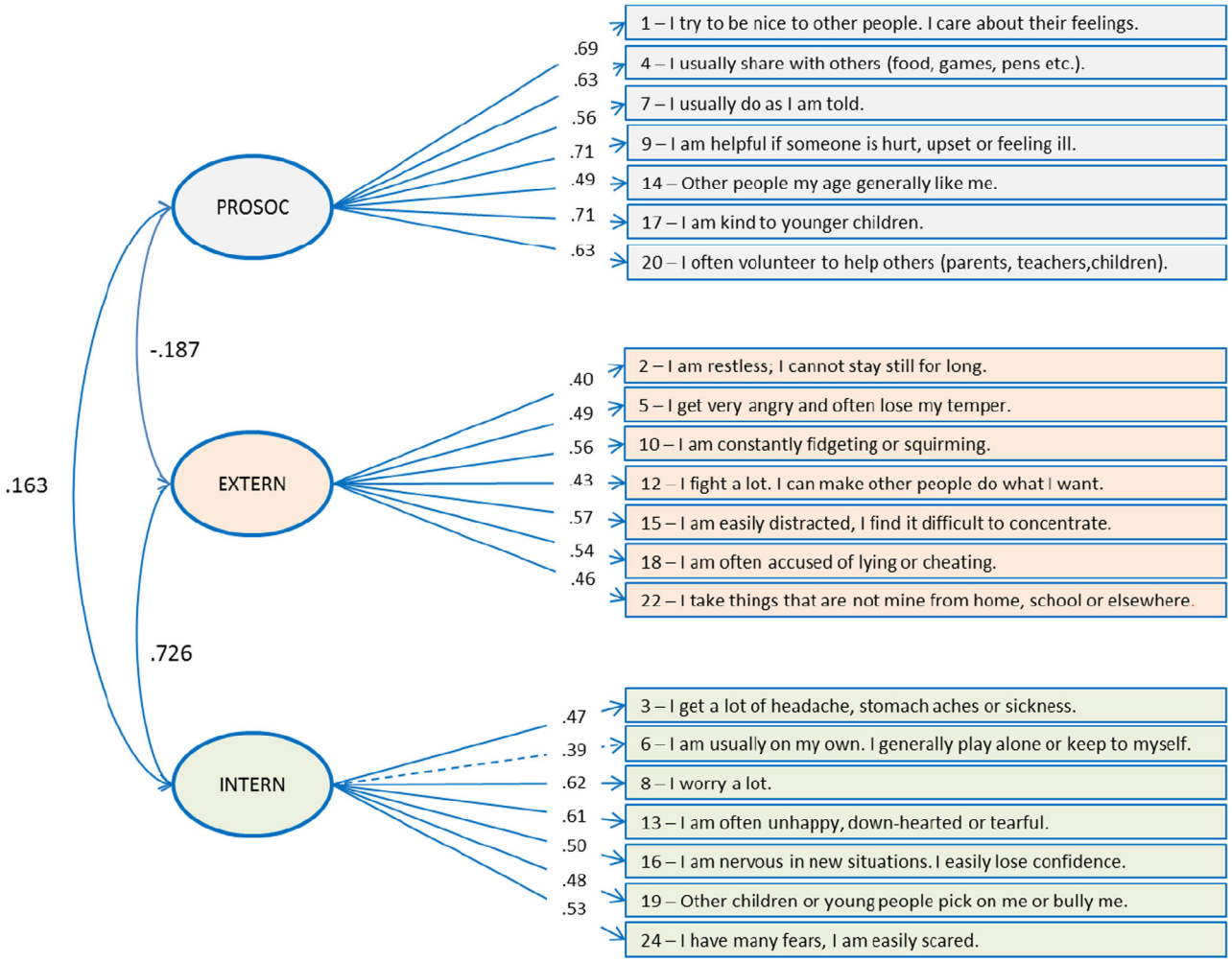

Fig. 2. Plain three-factor solution after exclusion of four items. WLSMV estimator with Mplus version 7. YX-standardized coefficients. N = 3,439; Clusters = 41; Chi-Square = 695.042, d.f. = 186, $p < 0.001$; RMSEA = 0.028; CFI = 0.911; TLI = 0.900.

DISCUSSION

Results from our confirmatory factor analyses of data collected among adolescents from the Western Cape, South Africa, indicate that a five-factor solution of the 25 items of the SDQ, as suggested by Goodman does not fit well (Goodman, 1997). After removing four items, two for conceptual reasons and two based on analysis of internal consistency, the remaining 21 items fitted well into three factors: Internalizing problems, Externalizing problems, and Prosocial behavior. Similar results have been reported in other studies (Di Riso *et al*., 2010; Dickey & Blumberg, 2004; El-Keshky & Emam, 2015; Gómez-Beneyto *et al*., 2013; Ortuño-Sierra *et al*., 2015; Ruchkin *et al*., 2008). The three factors showed strong measurement invariance across genders and language spoken at home (English, isiXhosa and Afrikaans).

According to Goodman and colleagues there is not one single best set of subscales to use in the SDQ (Goodman *et al*., 2010). The optimal choice may depend in part upon one's study population and study aims. If the number of factors, however, is permitted to vary freely across studies, comparing findings across studies will be almost impossible, or at least quite demanding. The search for a standard model with a specific number of factors and identical items covering each factor should therefore continue. As already mentioned, in this study based on data collected among secondary school students from the Western Cape, South Africa, a three-factor model obtained acceptable fit. Instead of a five-factor scale with five items in each factor, we ended up with three factors, covered by seven indicators each.

Model fit, which was already acceptable, could be improved by introducing three cross-loadings and one correlation between error terms (within one factor). These adjustments were considerably fewer than the number of adjustments needed to make the fit of the five-factor model reasonably good, and the plain three-factor model had much better fit than the plain five-factor model.

In spite of this empirical support for the three-factor model of the SDQ-S, there is room for improving psychometric properties of the three factors, particularly the Externalizing and the Internalizing factors. Improvements could potentially be obtained by changing some of the existing items. Hagquist (2007) suggested that items consisting of more than one statement could be split. The most obvious example of such an item is number 12: "I fight a lot. I can make other people do what I want." It is rather obvious that fighting a lot does not necessarily imply ability to persuade others. This item has obtained a low loading on its

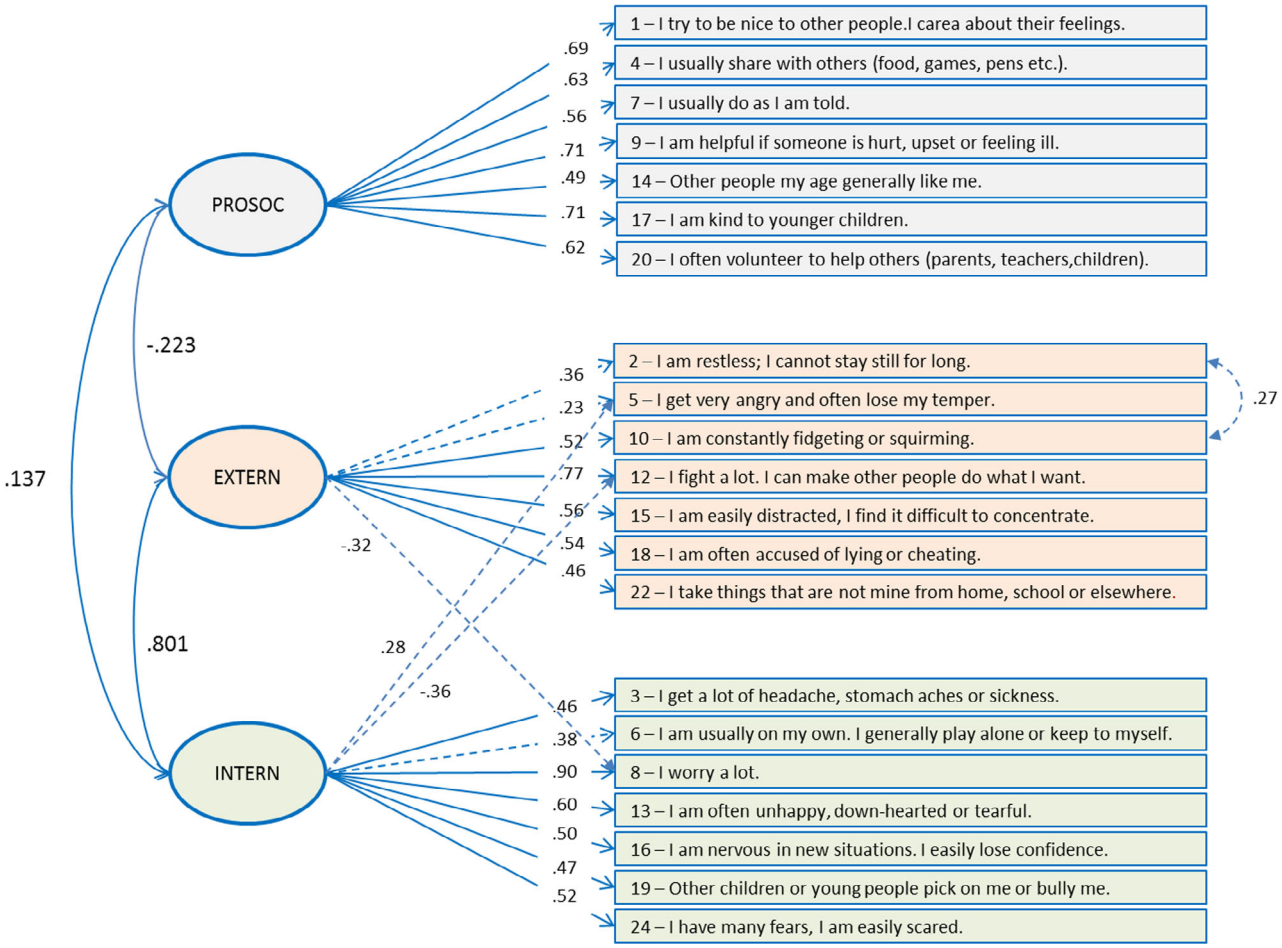

Fig. 3. Modified three-factor solution. WLSMV estimator with Mplus version 7. YX-standardized coefficients. $N = 3{,}439$; Clusters = 41; Chi-Square = 568.046, d.f. = 182, $p < 0.001$; RMSEA = 0.025 (0.023–0.027); CFI = 0.933; TLI = 0.922.

Table 2. *Testing measurement invariance across gender for clean 3-factor model. WLSMV estimator with Mplus version 7*

Model	Chi square	d.f.	P<	RMSEA	CFI	TLI
Configural (factor structure)	826.543	372	0.001	0.027	0.922	0.912
Metric (factor loadings) (weak invariance)	863.219	390	0.001	0.027	0.919	0.913
Scalar (intercepts) (strong invariance)	895.383	408	0.001	0.027	0.917	0.914
Metric vs. configural	77.159	18	0.001			
Scalar vs. configural	134.691	36	0.001			
Scalar against metric	70.342	18	0.001			

Table 3. *Testing measurement invariance[a] across gender for modified 3-factor model. WLSMV estimator with Mplus version 7*

Model	Chi square	d.f.	P<	RMSEA	CFI	TLI
Configural (factor structure)	725.771	364	0.001	0.024 (0.022–0.027)	0.938	0.929
Scalar (intercepts) (strong invariance)	802.104	403	0.001	0.024 (0.022–0.027)	0.932	0.929
Scalar vs. configural	136.096	39	0.001			

Note: [a]Measurement invariance testing with metric model and ordered polytomous categorical variables is not possible when a variable loads on more than one factor.

factor (Externalizing) in the plain three-factor model (0.43) and loadings on two different factors in the modified model. Avoiding items with more than one potential meaning would be a good principle when constructing scales.

Validation studies of the SDQ-S version suggests that the factorial structure of the five dimensions measured by the tool do not always yield the same structure across different ethnic or cultural groups (Stevanovic, Urbán, Atilola *et al.*, 2015). In our

Table 4. *Testing measurement invariance across language groups for clean 3-factor model. WLSMV estimator with Mplus version 7*

Model	Chi square	d.f.	P<	RMSEA	CFI	TLI
Configural (factor structure)	936.733	558	0.001	0.026	0.918	0.907
Metric (factor loadings) (weak invariance)	1004.960	594	0.001	0.026	0.911	0.905
Scalar (intercepts) (strong invariance)	1073.247	630	0.001	0.026	0.904	0.904
Metric vs. configural	142.317	36	0.001			
Scalar vs. configural	253.389	72	0.001			
Scalar against metric	144.808	36	0.001			

Table 5. *Testing measurement invariance[a] across language groups for modified 3-factor model. WLSMV estimator with Mplus version 7*

Model	Chi square	d.f.	P<	RMSEA	CFI	TLI
Configural (factor structure)	840.419	546	0.001	0.023 (0.020–0.026)	0.936	0.926
Scalar (intercepts) (strong invariance)	989.855	624	0.001	0.024 (0.021–0.027)	0.921	0.920
Scalar vs. configural	248.368	78	0.001			

Note: [a]Measurement invariance testing with metric model and ordered polytomous categorical variables is not possible when a variable loads on more than one factor.

Table 6. *Sum scores and corresponding latent variables correlated with sociodemographic factors. (For all analyses with latent variables n = 3,542)*

	Prosocial			Externalizing			Internalizing		
	Latent	Latent modified	Sumscore/ n	Latent	Latent modified	Sumscore/ n	Latent	Latent modified	Sumscore/ n
Ethnicity black (Black = 1; Other = 2)	−0.098***	−0.098**	−0.091*** 3,446	0.080*	0.061	0.078** 3,442	−0.012	0.012	0.008 3,443
Ethnicity white (White = 1; Other = 2)	0.078*	0.078*	0.063* 3,446	−0.020	−0.047	−0.010 3,442	0.002	0.011	0.002 3,443
Parents education (1 = high;2 = low/miss)	−0.161***	−0.161***	−0.145*** 3,542	0.080***	0.106***	0.066*** 3,542	0.058*	0.049*	0.048* 3,542
Doing well at school (n19rec) (1–5 – 5 = best)	0.166***	0.166***	0.143*** 3,449	−0.189***	−0.187***	−0.144*** 3,447	−0.105***	−0.126***	−0.075*** 3,448
Repeated school year (n20rec) (0 = no; 1 = yes)	−0.127***	−0.127***	−0.115*** 3,447	0.164***	0.171***	0.120*** 3,445	0.072*	0.086**	0.064** 3,444
Alcohol > one drink per day (n21rec) (0–2 – 2 = high)	−0.125***	−0.125***	−0.096*** 3,428	0.278***	0.264***	0.199*** 3,422	0.071***	0.105***	0.043* 3,425
Suicidal ideation (n189rec) (0 = no; 1 = yes)	−0.104***	−0.104***	−0.094*** 3,440	0.316***	0.324***	0.230*** 2,438	0.307***	0.317***	0.237*** 2,437
Victimized (1–5, 5 = Very often. Metric – MLR)	−0.132***	−0.132***	−0.123*** 3,432	0.326***	0.342***	0.241*** 3,431	0.248***	0.257***	0.205*** 3,430
Positive attitude towards school (1–5, 5 = pos) (Metric – MLR estimator)	−0.200***	−0.200***	−0.182*** 3,431	0.091***	0.111***	0.068*** 3,429	0.011	0.008	0.011 3,429
Negative attitude towards school (1–5, 5 = neg) (Metric – MLR estimator)	0.182***	0.182***	−0.157*** 3,433	−0.122***	−0.134***	−0.088*** 3,432	−0.021	−0.031	−0.034 3,431

Note: *$p < 0.05$,
**$p < 0.01$,
***$p < 0.001$.

study, strong measurement invariance was confirmed across genders as well as across languages (isiXhosa, English, Afrikaans) and for the plain as well as for the modified three factor models. This is important for comparison of findings between male and female school students as well as across groups defined by language spoken at home. In the present data from Cape Town, South Africa, such comparisons are valid for analyses of differences in latent SDQ subscale means as well as for comparisons of associations between latent SDQ factors and covariates across gender and language groups.

If every study applying the SDQ scale comes up with a different set of factors, this will of course limit the comparability

of findings across studies. Ideally, a good scale for the measurement of strengths and difficulties among adolescents should have a high level of measurement invariance across subgroups, study sites, countries and cultures. If analyses similar to the ones presented in this paper are carried out on data from other sites and countries, we could potentially learn more about the measurement invariance of the three-dimensional solution of the SDQ scale. Testing of modified and extended versions may be needed in order to obtain a tool with improved psychometric properties. Given that the SDQ is now used in studies in more than 80 countries, this is an important research task.

Latent variable analyses are not standard skills in all research environments, and many researchers would for many purposes, prefer to use additive indices or sum scores rather than latent variables. Comparisons of associations with covariates presented in this paper confirmed the usefulness also of this approach when analyzing SDQ data. Although associations tend to be weaker when using sum scores, the number of significances and the direction and strength of associations are similar.

The grouping of items by factors in the three-factor models presented in previous studies is remarkably consistent and correspond well with the results presented in this article. Furthermore, the four items that in our study have been excluded from the scale in order to make remaining items fit well with a three-factor solution have been shown to be problematic also in previous studies. Item number 11 ("I have one good friend or more") turned out to have low standardized factor loadings in studies from Italy (Di Riso *et al*., 2010) and Spain (Ortuño-Sierra *et al*., 2015). Item number 21 ("I think before I do things") showed low factor loading in the study from Italy (Di Riso *et al*., 2010). Item number 23 ("I get on better with adults than with people my age") had low loadings in studies from USA (Dickey & Blumberg, 2004; Ruchkin *et al*., 2008) and Spain (Di Riso *et al*., 2010; Ortuño-Sierra *et al*., 2015). Item number 25 ("I finish the work I'm doing. My attention is good") showed low standardized factor loadings in one study from Spain (Di Riso *et al*., 2010). If these items had been excluded from the three-factor models presented in previous studies, the consistency across studies might have been even higher.

SDQ data can be used for clinical purposes as well as in research. As Jum C. Nunnally wrote in his classic text on psychometric theory from 1967, what a satisfactory level of reliability is depends on how a measure is being used (Nunnally, 1967). He maintained that in early stages of research it might be sufficient to work with instruments having a reliability as low as 0.60 or even 0.50. By mathematically correcting for attenuation (less than perfect reliability), it is possible to estimate strengths of associations when reliability is higher or perfect. When scales are more well established and used in research based on data from for instance community studies, scales with reliability higher than 0.70 or even approaching 0.80 is preferred. When used in clinical settings, however, and when the purpose is to test individual clients or patients, a reliability of 0.90 is the minimum that should be tolerated, and according to Nunnally, a reliability of 0.95 "should be considered the desirable standard" (Nunnally, 1967).

This is reflected in more modern methods literature, for instance in Kattan and Cohen's (2009) *Encyclopedia of medical decision making*. Referring to Bland and Altman (1997), they state that in clinical settings, a value of alpha greater than 0.90 should be sought. In order to be used responsibly in clinical settings, the measurement of Internalizing problems, Externalizing problems, and Prosocial behavior needs to be improved and refined considerably over and above its present levels of internal consistency.

STUDY LIMITATIONS AND STRENGHTS

Among students invited to participate in the study, consents (also from parents for those younger than 18 years) were obtained from slightly more than half (55.3%). This might have been a critical problem if the purpose was to report estimates of population prevalence and means. Non-response is generally assumed to be less of a problem when the purpose of the study is examination of associations (Knudsen, Hotopf, Skogen, Øverland & Mykletun, 2010). Since disadvantaged groups tend to be over-represented among non-participants in data collections among adolescents, this may have reduced variance on SDQ indicators and thereby reduced strength of associations correspondingly. Provided that this had a uniform effect across indicators, reducing all intercorrelations to approximately the same degree, the dimensionality of the SDQ scale would remain rather unchanged. Gerrits, Van den Oord and Voogt (2001), in a study from the Netherlands have demonstrated that non-participation in studies on children's psychosocial adjustment does not necessarily introduce any bias at all. Still non-response has introduced some uncertainty with regard to the validity of our findings.

Collection of data in classroom contexts is an efficient way of collecting data. Data were collected by trained personnel, and teachers and other school personnel were not present during data collections. Large number of students in class was a challenge, but no major irregularities during data collections were reported. The questionnaire was given a youth-friendly design with colors, varying fonts, and exciting lay-out elements. This may have made the filling out less boring, more entertaining, and contributed to less missing on each questionnaire item.

Although weaknesses of the original five-factor model of SDQ have been shown in several previous studies, and support has been found for three-factor and four-factor solutions, no one has actually suggested an alternative to the five-factor model. The model suggested in the present publication is obviously not any kind of final answer. There is a need to further refine the measuring of the three dimensions in order to improve internal consistency and to examine psychometric qualities in new contexts.

CONCLUSIONS

The SDQ has achieved widespread use in a large number of countries across all continents. Consistent with several previous studies we conclude that in community samples the questionnaire functions better when analysed as three-dimensional (Externalizing problems, Internalizing problems, Prosocial behavior) rather than as consisting of five dimensions. In the present study, four items had to be removed in order for the scale to function well. Further research and improvement of the SDQ is warranted to settle the dimensionality issue, in order to improve its psychometric properties, and also in order to develop new versions with a high level of measurement invariance across sites, countries and cultures.

To the extent that dimensions like "Hyperactivity" and "Peer problems" are important to keep as separate aspects in clinical settings, new items should be developed and carefully tested in order to improve the internal consistency of these two dimensions.

ETHICS

The study was approved by the Human Research Ethics Committee, Faculty of Health Sciences, University of Cape Town (REC Ref: 268/2010), by the Western Cape Education Department and the Western Cape Department of Health, and by the Western Norway Regional Committee for Medical and Health Research Ethics. The authors attest that all procedures followed in the research study complied with the ethical standards and guidelines of the ethics committees. All participants and their parents were informed about the nature of the study as well as its aims. As participants were under the age of 18, parental consent as well as personal assent was needed before participation in the study. The present publication is based on baseline data from a cluster-randomized controlled trial carried out in Cape Town, South Africa. Trial registration number: Controlled Trials ISRCTN56270821.

ACKNOWLEDGEMENTS

The full title of the project is: 'Promoting sexual and reproductive health among adolescents in southern and eastern Africa – mobilising schools, parents and communities.' Acronym: PREPARE. The partners and principal investigators include: University of Cape Town (Cathy Mathews), Muhimbili University of Health and Allied Sciences (Sylvia Kaaya), University of Limpopo (Hans Onya), Makerere University (Anne Katahoire), Maastricht University (Hein de Vries), University of Exeter (Charles Abraham), University of Oslo (Knut-Inge Klepp), University of Bergen (Leif Edvard Aarø – coordinator). We would like to express our gratitude to the members of the PREPARE Scientific Advisory Committee: Nancy Darling, Oberlin College, Ohio, USA; Jane Ferguson, World Health Organization, Geneva, Switzerland; Eleanor Maticka-Tyndale, University of Windsor, Canada, and David Ross, London School of Hygiene and Tropical Medicine, UK. See also the project homepage http://prepare.b.uib.no/. We would like to express appreciation to our colleagues in the Western Cape Department of Health, the City of Cape Town Health Department, and the Desmond Tutu HIV Foundation and the Western Cape Department of Education for supporting the research, particularly the school nurses and health promoters, Tracy Naledi, Michelle Williams, Karen Jennings, Linda-Gail Bekker, Thereza Bothma, Estelle Lawrence, Anik Gevers and Estelle Matthews. We appreciate all the principals, teachers and education officials who facilitated the PREPARE trial. We acknowledge the work the research team including Petra De Koker, Joy Koech, Tracy McClinton Appollis, Mariette Momberg, Sandra de Jager, Karin Webber, and Leah Demetri and all the intervention facilitators and data collectors without whom this research would not have been possible.

FUNDING INFORMATION

The PREPARE study was funded by the EC Health Research Programme (Under the 7th Framework Programme). Grant Agreement number: 241945.

DATA AVAILABILITY STATEMENT

Data used in the PREPARE study are not available for sharing due to ethical and data management requirements. The researchers are open to collaboration.

REFERENCES

Aarø, L.E., Mathews, C., Kaaya, S., Katahoire, A.R., Onya, H., Abraham, C. *et al*. (2014). Promoting sexual and reproductive health among adolescents in southern and eastern Africa (PREPARE): Project design and conceptual framework. *BMC Public Health*, *14*, 54.

Becker, A., Hagenberg, N., Roessner, V., Woerner, W. & Rothenberger, A. (2004). Evaluation of the self-reported SDQ in a clinical setting: Do self-reports tell us more than ratings by adult informants? *European Child & Adolescent Psychiatry*, *13*, 17–24.

Bland, J.M. & Altman, D.G. (1997). Statistics notes: Cronbach's alpha. *British Medical Journal*, *314*, 572.

de Vries, P.J., Davids, E.L., Mathews, C. & Aarø, L.E. (2018). Measuring adolescent mental health around the globe: Psychometric properties of the self-report strengths and difficulties questionnaire in South Africa, and comparison with UK, Australian and Chinese data. *Epidemiology and Psychiatric Sciences*, *27*, 369–380.

Di Riso, D., Salcuni, S., Chessa, D., Raudino, A., Lis, A. & Altoe, G. (2010). The strengths and difficulties questionnaire (SDQ): Early evidence of its reliability and validity in a community sample of Italian children. *Personality and Individual Diffferences*, *49*, 570–575.

Dickey, W.C. & Blumberg, S.J. (2004). Revisiting the factor structure of the strengths and difficulties questionnaire: United States. *Journal of the American Academy of Child & Adolescent Psychiatry*, *43*, 1159–1167.

El-Keshky, M. & Emam, M. (2015). Emotional and behavioral difficulties in children referred for learning disabilities from two Arab countries: A cross-cultural examination of the strengths and difficulties questionnaire. *Research in Developmental Disabilities*, *36*, 459–469.

Garrido, L.E., Barrada, J.R., Agusvivas, J.A., Martínez-Molina, A., Arias, V.B., Golino, H.F. *et al*. (2018). Is small still beautiful for the strengths and difficulties questionnaire? Novel findings using exploratory structural equation modeling. *Assessment*, *27*, 1349–1367. https://doi.org/10.1177/1073191118780461.

Gerrits, M.H., Van den Oord, E.J.C.G. & Voogt, R. (2001). An evaluation of nonresponse bias in peer, self, and teacher ratings of children's psychosocial adjustment. *Journal of Child Psychology and Psychiatry*, *42*, 593–602.

Gómez-Beneyto, M., Nolasco, A., Moncho, J., Pereyra-Zamora, P., Tamayo-Fonesca, N., Munarriz, M. *et al*. (2013). Psychometric behavior of the strengths and difficulties questionnnaire (SDQ) in the Spanish national health survey 2006. *BMC Psychiatry*, *13*, 95.

Goodman, A., Lamping, D.L. & Ploubidis, G.B. (2010). When to use broader internalizing and externalizing subscales instead of the hypothesized five subscales on the Strenghts and difficulties questionnaire (SDQ): Data from British parents, teachers and children. *Journal of Abnorm Child Psychology*, *38*, 1179–1191.

Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, *38*, 581–586.

Hagquist, C. (2007). The psychometric properties of the self-reported SDQ – an analysis of Swedish data based on the Rasch model. *Personality and Individual Differences*, *43*, 1289–1301.

Hoosen, N., Davids, E.L., de Vries, P.J. & Shung-King, M. (2018). The strengths and difficulties questionnaire (SDQ) in Africa: A scoping review of its application and validation. *Child and Adolescent Psychiatry and Mental Health*, *12*, 6.

Kattan, M.W. & Cohen, M.E. (2009). *Encyclopedia of medical decision making*, Vol. *1*. Thousand Oaks, CA: Sage.

Kessler, R.C., Amminger, G.P., Aguilar-Gaxiola, S., Alonso, J., Lee, S. & Ustun, T.B. (2007). Age of onset of mental disorders: A review of recent literature. *Current Opinion in Psychiatry*, *20*, 359–364.

Klasen, H. & Crombag, A. (2013). What works where? A systematic review of child and adolescent mental health interventions for low and middle income countries. *Social Psychiatry and Psychiatric Epidemiology*, *48*, 595–611.

Knudsen, A.K., Hotopf, M., Skogen, J.C., Øverland, S. & Mykletun, A. (2010). The health status of nonparticipants in a population-based health study: The Hordaland health study. *American Journal of Epidemiology*, *172*, 1306–1314.

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

Lu, C., Black, M.M. & Richter, L.M. (2016). Risk of poor development in young children in low-income and middle-income countries: An estimation and analysis at the global, regional and country level. *The Lancet Global Health*, *4*, e916–e922.

Lu, C., Li, Z. & Patel, V. (2018). Global child and adolescent mental health: The orphan of development assistance for health. *PLoS Medicine*, *15*, e1002524.

Lund, C., Kleintjies, S., Kakuma, R., Flisher, A.J. & MHaPP Research Programme Consortium. (2010). Public sector mental health systems in South Africa: Inter-provincial comparisons and policy implications. *Social Psychiatry and Psychiatric Epidemiology*, *45*, 393–404.

Mathews, C., Eggers, S., Townsend, L., Aarø, L.E., de Vries, P.J. *et al*. (2016). Effects of PREPARE, a multi-component, school-based HIV and intimate partner violence (IPV) prevention programme on adolescent sexual risk behaviour and IPV: Cluster-randomized controlled trial. *AIDS and Behavior*, *20*, 1821–1840.

Mellins, C.A., Xu, Q., Nestadt, D.F., Knox, J., Kauchali, S., Arpadi, S. *et al*. (2018). Screening for mental health among young south African children: The use of the strengths and difficulties questionnaire. *Global Social Welfare*, *5*, 29–38.

Nunnally, J.C. (1967). *Psychometric theory*. New York: McGraw-Hill.

Ortuño-Sierra, J., Fonseca-Pedrero, E., Paino, M., Sastre i Riba, S. & Muñiz, J. (2015). Screening mental health problems during adolescence: Psychometric properties of the Spanish version of the strengths and difficulties questionnaire. *Journal of Adolescence*, *38*, 49–56.

Patel, V., Flisher, A.J., Nikapota, A. & Malhotra, S. (2008). Promoting child and adolescent mental health in low and middle income countries. *Journal of Child Psychology and Psychiatry*, *49*, 313–334.

Percy, A., McCrystal, P. & Higgins, K. (2008). Confirmatory factor analysis of the adolescent self-report strengths and difficulties questionnaire. *European Journal of Psychological Assessment*, *24*, 43–48.

Rathod, S., Pinninti, N., Irfan, M., Gorczynski, P., Rathod, P., Gega, L. *et al*. (2017). Mental health service provision in low- and middle-income countries. *Health Services Insights*, *10*, 1–7.

Rønning, J.A., Handegaard, B.H., Sourander, A. & Mørch, W.-T. (2004). The strengths and difficulties self-report questionnaire as a screening instrument in Norwegian community samples. *European Child & Adolescent Psychiatry*, *13*, 73–82.

Ruchkin, V., Jones, S., Vermeiren, R. & Schwab-Stone, M. (2008). The strengths and difficulties questionnaire: The self-report version in American urban and suburban youth. *Psychological Assessment*, *20*, 175–182.

Stevanovic, D., Urbán, R., Atilola, O., Vostanis, P., Singh Balhara, Y.P., Avicenna, M. *et al*. (2015). Does the strengths and difficulties questionnaire – Self report yield invariant measurements across different nations? Data from the international child mental health study group. *Epidemiology and Psychiatric Sciences*, *24*, 323–334.

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*, 127–144.

Van Roy, B., Veenstra, M. & Clench-Aas, J. (2008). Construct validity of the five-factor strengths and difficulties questionnaire (SDQ) in pre-, early, and late adolescence. *Journal of Child Psychology and Psychiatry*, *49*, 1304–1312.

Received 26 July 2021, Revised 23 February 2022, accepted 9 March 2022