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

P. J. de Vries[1,2]*, E. L. Davids[1,2], C. Mathews[2,3] and L. E. Aarø[4]

[1] *Division of Child & Adolescent Psychiatry, University of Cape Town, South Africa*

[2] *Adolescent Health Research Unit, University of Cape Town, South Africa*

[3] *Health Systems Research Unit, Medical Research Council; & School of Public Health and Family Medicine, University of Cape Town, South Africa*

[4] *Department of Health Promotion, Norwegian Institute of Public Health, Bergen, Norway*

Aims. This study evaluated the psychometric properties of the Strengths and Difficulties Questionnaire Self-Report (SDQ-S) in South African adolescents, and compared findings with data from the UK, Australia and China.

Methods. A sample of 3451 South African adolescents in grade 8, the first year of secondary school ($M_{age} = 13.7$ years), completed the SDQ-S in Afrikaans, English or isiXhosa. Means, group differences and internal consistency were analysed using SPSS V22, and confirmatory factor analyses were conducted using MPlus V7.

Results. In the South African sample, significant gender differences were found for four of the five sub-scale means and for total difficulties, but gender differences of alpha scores were negligible. The internal consistency for the total difficulties, prosocial behaviour and emotional symptoms sub-scales were fair. UK cut-off values for caseness (set to identify the top 10% of scores in a UK sample) led to a higher proportion of South African adolescents classified in the 'abnormal' range on emotional and peer difficulties and a lower proportion classified in the 'abnormal' range for hyperactivity. South African cut-offs were therefore generated. The cross-country comparison with UK, Australian and Chinese data showed that South African adolescent boys and girls had the highest mean scores on total difficulties as well as on the subscales of emotional symptoms and conduct problems. In contrast, South African boys and girls had the lowest mean scores for hyperactivity/inattention. The UK boys and girls had the highest mean scores for hyperactivity/inattention, while the Australian sample had the highest scores for prosocial behaviours. The Chinese boys had the highest peer problem mean scores and Chinese boys and girls had the lowest means on prosocial behaviours. Confirmatory factor analyses showed significant item loadings with loadings higher than 0.40 for the emotional and prosocial behaviour sub-scales on the five-factor model, but not for all relevant items on the other three domains.

Conclusions. Findings support the potential usefulness of the SDQ-S in a South African setting, but suggest that the SDQ-S should not be used with UK cut-off values, and indicate the need for further validation and standardisation work in South African adolescents. We recommend that in-country cut-offs for 'caseness' should be used for clinical purposes in South Africa, that cross-country comparisons should be made with caution, and that further examination of naturalistic clusters and factors of the SDQ should be performed in culturally and contextually diverse settings.

Introduction

Standardised measurement of child and adolescent mental health has become increasingly important as a way to identify young people at risk, to define and monitor outcomes, to predict clinical service needs, and for comparative global studies about mental health (Angold & Costello, 2009; Deighton *et al.* 2014). There is also a growing interest in the voice of adolescents as part of assessments, such as through self-report measures (Deighton *et al.* 2014). For large-scale implementation, freely available, well-standardised instruments are likely to be most feasible. A recent review of self-report measures for potential implementation in a UK child and adolescent mental health setting identified 11 potentially useful instruments (Deighton *et al.* 2014). The authors commented

---

* Address for correspondence: Prof P. J. de Vries, Division of Child & Adolescent Psychiatry, University of Cape Town, 46 Sawkins Road, Rondebosch 7700, South Africa.

(Email: petrus.devries@uct.ac.za)

that, whilst all these instruments had positive aspects, none had sufficient psychometric evidence to demonstrate that they could measure severity and change over time. They suggested further work to examine the feasibility and psychometric credibility of such instruments for policy and practice (Deighton et al. 2014).

Access to mental health services in low- and middle-income countries (LMICs) is greatly impacted by the gap between need for and access to mental health services, where the treatment gap, particularly for children and adolescents, can be as high as 90% (Alem et al. 2009; Kieling et al. 2011). Implementation of standardised self-report measures may therefore represent a powerful strategy to identify young people at risk of mental health problems, to monitor outcomes and to inform policy development. An important aspect of global service and policy development may also be cross-country comparisons to examine the similarities and differences in rates of mental health problems within or between geographical regions. Measurement instruments that can perform such comparisons may therefore become invaluable as tools in global implementation science research.

The Strengths and Difficulties Questionnaire (SDQ) is a screening tool that has been translated into many languages, has been implemented in mental health settings in many countries around the globe, and is considered a valid, rapid measure for emotional and behavioural problems in a number of settings (Goodman et al. 2000a, b; Skinner et al. 2014). It is a 25-item measure and consists of five sub-scales: emotional symptoms, conduct problems, hyperactivity/inattention, peer problems, and prosocial behaviour. The composite score of the first four sub-scales represents a 'total difficulties' score (the higher the total difficulties score the more significant problems), while the latter sub-scale reflects prosocial behaviour (the higher the score the better the prosocial behaviour) (Kersten et al. 2016). Three versions of the SDQ are available, a parent form (SDQ-P), a teacher form (SDQ-T) and a self-report form (SDQ-S). Here we focus on the SDQ-S.

The SDQ uses cut-off scores for 'caseness' to define scores as normal (80% of scores), borderline (80th–90th percentile scores) or abnormal (>90th percentile scores). These cut-off values were derived from UK data (Kremer et al. 2015). The use of UK cut-off scores may not always be suitable when the SDQ is used in different countries and contexts (Goodman, 1997). Revised cut-off values have been calculated for a number of countries, including China (Du et al. 2008), India (Bhola et al. 2016) and the Democratic Republic of Congo (Kashala et al. 2005). Apart from using the SDQ to identify 'caseness', Goodman and colleagues have suggested that SDQ population means could

also be useful as an indicator of the prevalence of mental disorders (Goodman & Goodman, 2011). They confirmed this using a large-scale UK sample of SDQ and Development and Well-Being Assessment (DAWBA) data (Goodman & Goodman, 2011). However, cross-country differences in SDQ population means were not supported by the DAWBA individual clinical assessment of study participants. The authors of the SDQ therefore suggested that differences in cross-country means may not be indicators of differential rates of mental disorders (Goodman & Goodman, 2011), but may rather reflect the lack of population-specific norms to make 'valid' cross-country comparisons (Goodman et al. 2012).

A recent scoping review of 36 studies evaluating the use of the SDQ in Africa suggested the SDQ was largely used to assess mental health of children and adolescents who were either affected or infected by HIV/AIDS. In addition to this, the review highlighted: (i) few evaluations of the psychometric properties of the SDQ on the African continent; (ii) the use of some but not all subscales of the SDQ; and (iii) the presentation of clinical cut-offs specific to an African country in only one study. When considering this, however, even though a number of studies in South Africa have used the SDQ, there has to date not been any evaluation of the psychometric properties of the SDQ, and it remains unknown whether the UK cut-off values would be appropriate in a South African setting.

To our knowledge, there has been only one study that examined SDQ-S population means in a sub-Saharan Africa country. In this study the SDQ-S was used among 2367 adolescents from India, Indonesia, Nigeria, Serbia, Turkey, Bulgaria and Croatia (Stevanovic et al. 2014). The findings showed that the five-factor model of the SDQ-S did not fit any of these countries, and indicated that different models fitted different countries. Stevanovic et al. (2014) therefore provided further support that the SDQ-S may not be suitable for cross-country comparison, but that appropriate norms could be valuable within-country. Interestingly, the psychometric properties of the SDQ-S in China, an upper-middle income country like South Africa, were shown to be good with high reliability and validity (Du et al. 2008). Similarly, findings in an Australian study found good psychometric properties for the SDQ-S (Mellor, 2004).

Given that no previous psychometric evaluation of the SDQ has taken place in South Africa, the current study therefore aimed to: (i) examine the means, standard deviation, distribution and internal consistency of the SDQ-S in a representative sample of adolescent boys and girls; (ii) compare gender differences in the above psychometric properties; (iii) determine the

proportion of boys and girls who scored in the 'normal', 'borderline' and 'abnormal range' of SDQ-S subscales based on UK norms, and, if required, to generate South African cut-off scores; (iv) compare mean SDQ-S scores of the South African sample to previously reported normative data for SDQ-S scores in UK, Australian and Chinese samples; and (v) determine whether the South African SDQ-S data would fit the five-factor structure of the original UK SDQ-S. We hypothesised that significant differences on the basis of gender will exist in SDQ-S scores as found in previous studies (Becker *et al.* 2015; Kremer *et al.* 2015), hypothesised that South African cut-off scores may be required, as reported in other studies from LMICs (Kashala *et al.* 2005; Menon *et al.* 2007; Bakare *et al.* 2010; Cortina *et al.* 2013), and that there would be only partial support for the five-factor structure, as reported elsewhere (Rønning *et al.* 2004; Richter *et al.* 2011; Stevanovic *et al.* 2014). We did not hypothesise any specific patterns of similarities or differences between the South African, UK, Australian and Chinese data, but were keen to explore the cross-country potential of the instrument.

Methods

Participants

This was a sub-study of a larger cluster, randomised controlled trial, PREPARE (Promoting Sexual and Reproductive Health among Adolescents in Southern and Eastern Africa) (Aarø *et al.* 2014). We randomly sampled 42 public high schools from the database of high schools in the Western Cape Province in South Africa, and one school subsequently dropped out. Participants were Grade 8 adolescents (average age 13 years) attending these schools. For detail of the PREPARE trial methodology; see Aarø *et al.* (2014) and Mathews *et al.* (2016). We invited 6244 students to participate in the PREPARE trial and 3451 (55.3%) returned signed parental/care-giver 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 PREPARE study surveyed participants through a paper questionnaire at baseline, 6 and 12 months. All questions were provided in the three languages commonly spoken in the Western Cape Province (Afrikaans, English and isiXhosa). The multi-language questionnaire presented questions simultaneously in Afrikaans, English and isiXhosa. The questionnaires 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 are based on baseline data.

We used the English version of the SDQ-S and undertook standard procedures for translation and back-translation as required by the authors and developers of the SDQ to develop Afrikaans and isiXhosa versions. 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. When there were uncertainties, we contacted the original developer of the SDQ for clarification.

Procedure

The baseline survey, which included the SDQ-S, was conducted in classrooms during school hours, and completion of the questionnaire took on average 45 min.

Ethics approval

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, the Western Cape Department of Health, and by the Western Norway Regional Committee for Medical and Health Research Ethics.

Data analysis

In order to ensure comparability with previous SDQ studies, it was decided to base descriptive statistical analyses on the five-factor model (emotional, conduct, hyperactivity, peer, prosocial) without modifications. Descriptive analyses were carried out with IBM SPSS Statistics 22 and included calculation of means with confidence intervals and significance testing of differences across groups, standard deviations, and Cronbach's alpha coefficients. Confidence intervals and *p*-values were adjusted for cluster effects (individual learners within schools). The proportion of South African adolescents in the 'normal', 'borderline' and 'abnormal' range were calculated, based on UK norms. South African cut-offs were generated to classify the 80th–90th percentile band and >90th percentile bands. Mean scores on the total SDQ-S scale as well as

for subscales were compared with similar scores from samples in three other countries. A Chinese sample was selected among school students from 12 out of the 19 administrative districts in Shanghai (Du *et al.* 2008), an Australian sample was from schools across Victoria (Mellor, 2005), and UK data were from a nation-wide, representative sample of adolescents (Meltzer *et al.* 2000). Given that sufficient statistical information to control for design effects such as the cluster effect was not available for the other studies, confidence intervals for these data were based on the simple random sample assumption.

Analyses of dimensionality of the SDQ-S scale – confirmatory factor analyses with no restriction on inter-factor correlations and control for cluster effects – were carried out with Mplus version 7. All indicators were defined as (ordered) categorical and we used the Weighted Least Squares Mean and Variance adjusted (WLSMV) estimator. 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).

Results

Overall, 3451 adolescents participated in the baseline survey. Among the 3360 students who reported their gender at baseline, 39.7% (1334) were males and 60.3% (2026) were females. Most of the students (99.1%) were in the age range 12–16 years. The mean age was 13.7 years. Descriptive statistics (means and standard deviations) and Cronbach alpha values for SDQ-S subscales, total scale and impact are shown in Table 1. For two of the subscales (emotional symptoms, prosocial behaviour) and for total difficulties and impact, alpha values were 0.60 or higher for both genders combined. For the hyperactivity/inattention subscale, alpha was 0.52. The conduct problems and peer problems subscales obtained low alpha values of 0.37 and 0.29, respectively. Gender differences in alpha values were small.

Gender differences in mean scores were tested with adjustment for age (Table 2). Girls scored significantly higher than boys on emotional symptoms (diff = 1.16; $p < 0.001$), prosocial behaviour (diff = 0.63; $p < 0.001$) and total difficulties (diff = 0.56; $p < 0.01$). Boys scored significantly higher than girls on conduct problems (diff = 0.30; $p < 0.001$) and hyperactivity/inattention (diff = 0.21; $p < 0.01$).

To explore the distribution of 'caseness' of SDQ-S data, SDQ-S scores were grouped (banded) into the three categories as presented by Goodman *et al.* (2000*a*). Table 3 shows the three-group banding of scores ('normal', 'borderline' and 'abnormal') for our sample alongside the UK norms (www.sdqinfo.com). The groupings were reflected in two ways. First, the

frequency of South African and UK adolescents above each of the cut-offs were presented. The UK norms were derived to show an 80–10–10 per cent distribution. Since the underlying initial scores are already forming discrete categories, only rough approximations were possible. When the same norms (UK) were applied to both samples, a few pronounced differences could be observed. In South Africa, there was a higher proportion with high scores on total difficulties (14.9 *v.* 9.2%), emotional symptoms (26.0 *v.* 11.2%), peer problems (33.7 *v.* 9.2%) and impact (16.9 *v.* 5.8%). The South African data had more favourable results for prosocial behaviour (15.7% with high scores in South Africa *v.* 8.6% in the UK sample). The proportion with high scores on hyperactivity/inattention was lower in the South African sample (3.7 *v.* 11.5%). We therefore also presented revised cut-off values to represent the South African data in an 80–10–10 distribution (see Table 3).

Mean scores on all SDQ-S scales by gender are shown for samples from four countries (South Africa (this study), UK (Meltzer *et al.* 2000), Australia (Mellor, 2005) and China (Du *et al.* 2008) in Table 4. All comparisons across samples were performed for boys and girls separately. The South African sample had higher mean scores than any of the other samples (and no overlap of confidence intervals) on emotional symptoms (both genders) and conduct problems (both genders, but slight confidence interval overlap with the Chinese sample among boys). On the peer problems subscale, the South African sample had higher scores than the UK and Australia (both genders), but lower than Chinese boys. The South African sample had lower mean scores on hyperactivity/inattention than adolescents in any of the other samples (slight overlap between confidence intervals between South Africa and Australian girls). With regard to prosocial behaviour, the South African sample was not much different from the other samples, except for Chinese boys, who had lower mean scores than any other group, irrespective of gender. On the total difficulties scale, South African and Chinese boys had the highest mean scores. On this scale South African girls had higher scores than girls from any of the other three countries. Impact scores could only be compared with UK data. The South African adolescents of both genders had mean scores higher than those in the UK sample on this scale.

The dimensionality of the SDQ-S scale was tested with confirmatory factor analyses. A 'clean' five-factor solution (with 5 SDQ-S items each loading onto one of 5 factors) obtained poor fit ($\chi^2 = 2710.14$; d.f. = 265; $p < 0.001$; RMSEA = 0.052; CFI = 0.629; TLI = 0.580). After a series of step-by-step modifications, based on modification indices and allowing significant cross-

Table 1. *SDQ means, standard deviations and internal consistency by gender (range of all scales 0–10, except 'Difficulties total' which ranges from 0 to 40)*

	All[a]				Boys				Girls			
	N	Mean	S.D.	Alpha	n	Mean	S.D.	Alpha	n	Mean	S.D.	Alpha
Emotional symptoms	3252	3.94	2.44	0.60	1278	3.26	2.24	0.55	1974	4.38	2.47	0.60
Conduct problems	3248	2.41	1.78	0.37	1279	2.61	1.85	0.38	1969	2.28	1.72	0.35
Hyperactivity/inattention	3255	2.73	2.01	0.52	1278	2.87	2.08	0.52	1977	2.64	1.96	0.52
Peer problems	3253	2.76	1.90	0.29	1279	2.86	1.87	0.26	1974	2.69	1.90	0.31
Prosocial behaviour	3270	7.76	2.12	0.68	1284	7.36	2.20	0.67	1986	8.02	2.02	0.68
Difficulties total	3238	11.84	5.49	0.69	1273	11.59	5.43	0.69	1965	12.00	5.52	0.70
Impact	3175	0.71	1.62	0.74	1242	0.70	1.61	0.74	1933	0.71	1.63	0.74

[a]Cases with missing on gender ($n = 14$) not included.

loading in the model, acceptable fit was obtained ($\chi^2 = 637.54$; d.f. = 258; $p < 0.001$; RMSEA = 0.021; CFI = 0.942; TLI = 0.933) (Fig. 1). Only two factors, emotional problems and prosocial behaviour, proved to have loadings higher than 0.40 on all five relevant indicators. Hyperactivity and conduct problems loaded on three relevant indicators each. Peer problems loaded on two relevant indicators only. Emotional problems loaded on one irrelevant indicator, but the coefficient was rather small (0.28). Prosocial behaviour loaded on six irrelevant indicators (four loadings larger than 0.40). The highest correlations between latent factors were between emotional problems and peer problems (0.81), between hyperactivity/inattention and conduct problems (0.73) and between emotional problems and hyperactivity/inattention (0.67).

Discussion

Measurement of child and adolescent mental health has become a growing aim for clinical, service development, and global research purposes. A number of self-report instruments are currently in use, all with their own areas of strengths and limitations. The SDQ has been used widely, but relatively little is known about the psychometric properties of the SDQ outside the UK. We therefore set out to explore the psychometric properties of the SDQ-S in a representative sample of South African adolescents in order to evaluate means, standard deviations and internal consistency of the sub-scales across gender, to examine the cut-off scores for 'caseness', to compare South African adolescent findings with data from three other countries, and to test out the five-factor model of the SDQ-S in our setting.

Gender and 'caseness' differences

In line with previous studies, the South African findings showed gender-based differences on SDQ-S sub-scale scores. In examining the parent, teacher and self-report versions of the SDQ gender differences were found in a Dutch sample (van Widenfelt *et al.* 2003) where boys had higher mean scores for conduct problem and girls had higher means for prosocial behaviour, similar to our findings. Similar gender differences were also reported in an Italian study (Di Riso *et al.* 2010). The

Table 2. *SDQ scales by gender, adjusted for age. Confidence intervals adjusted for cluster effects*

	Boys			Girls			Difference		
	n	Mean[a]	CI$_{95\%}$	n	Mean[a]	CI$_{95\%}$	Wald F	d.f.	P
Emotional symptoms	1278	3.24	3.13–3.35	1974	4.40	4.24–4.55	180.36	1, 40	0.000
Conduct problems	1279	2.59	2.47–2.72	1969	2.29	2.17–2.42	24.64	1, 40	0.000
Hyperactivity/inattention	1278	2.86	2.70–3.02	1977	2.65	2.51–2.79	9.65	1, 40	0.003
Peer problems	1279	2.82	2.66–2.98	1974	2.72	2.56–2.87	2.48	1, 40	0.123
Prosocial behaviour	1284	7.38	7.22–7.53	1986	8.01	7.89–8.14	53.27	1, 40	0.000
Total difficulties	1273	11.50	11.21–11.79	1965	12.06	11.67–12.44	7.76	1, 40	0.008
Impact	1242	0.68	0.58–0.78	1933	0.73	0.65–0.81	0.76	1, 40	0.387

[a]Estimated marginal means after adjustment for age.

Table 3. *Banding of SDQ raw scores in South Africa and UK.*

	U.K.[a] Self-report UK norms N = 4228 (11–15 years)		South Africa[b] Self-report UK norms N = 3250 (12–16 years)		South Africa[b] Self-report SA norms N = 3250 (12–16 years)	
	Band	%	Band	%	Band	%
Total difficulties	0–14	79.2	0–14	69.5	0–16	80.8
	15–17	11.6	15–17	15.6	17–19	10.1
	18–40	9.2	18–40	14.9	20–40	9.1
Emotional symptoms	0–4	78.9	0–4	61.3	0–6	84.3
	5	9.9	5	12.7	7	6.0
	6–10	11.2	6–10	26.0	8–10	9.7
Conduct problems	0–3	79.0	0–3	74.1	0–3	74.1
	4	10.4	4	13.5	4	13.5
	5–10	10.6	5–10	12.4	5–10	12.4
Hyperactivity/inattention	0–5	79.0	0–5	90.7	0–4	80.3
	6	9.5	6	5.6	5	10.4
	7–10	11.5	7–10	3.7	6–10	9.3
Peer problems	0–2	80.3	0–2	49.8	0–4	82.1
	3	10.5	3	16.5	5	8.9
	4–10	9.2	4–10	33.7	6–10	9.0
Prosocial behaviour	7–10	79.2	7–10	73.5	6–10	84.3
	6	12.2	6	10.8	5	8.0
	0–5	8.6	0–5	15.7	0–4	7.7
Impact	0	86.8	0	75.8	0	83.1
	1	7.4	1	7.3	1–2	5.9
	2–10	5.8	2–10	16.9	3–10	11.0

[a]http://www.sdqinfo.com/norms/UKSchoolNorm.html
[b]Present study.

gender differences seen in the current study, where girls scored higher on emotional symptoms and boys higher on conduct problems, may be explained by the fact that girls are at greater risk of internalised mental health problems (emotional symptoms) while boys are at greater risks of externalised mental health problems (conduct problems) (Needham & Hill, 2010). As outlined below, it will, however, be important to determine whether the statistically-significant gender differences observed are also clinically-significant.

Application of the UK cut-off values for 'caseness' to the South African sample led to extremely high rates of abnormal scores for emotional symptoms, peer problems and on the impact score. Using UK cut-off values, 26 and 34% of South African adolescents were rated in the 'abnormal' range for emotional and peer problems, in contrast to the normative expectation of 10%. There may therefore be a temptation to interpret these findings as indicative of higher mental health problems and lead to *post-hoc* hypotheses that South African adolescents may have a particular predisposition to mental health problems as a result of various genetic or environmental factors such as

poverty or cultural factors (Kashala *et al.* 2005; Myer *et al.* 2009). However, as we explored in this study, it is also possible that these results are attributable to measurement error. Goodman *et al.* (2012) very astutely suggested that the estimation of mental health disorders can be difficult across countries, particularly when population-specific norms are absent.

We are very mindful of the need to distinguish between 'statistically-significant' differences and 'clinically-significant' differences. All the differences observed and reported here were statistical ones, rather than based on any clinical criteria or clinical evaluation. We would certainly recommend that, before any conclusions are drawn, SDQ-S scores should be compared to appropriate, acceptable and standardised 'gold standard' clinical evaluations. An ideal next step would be to perform population-based studies in South Africa that might then allow cross-country and cultural comparison (Goodman *et al.* 2012). A South Africa study combining SDQ-S and standardised clinical evaluations in clinical and community samples will provide data to establish the sensitivity, specificity and reliability of SDQ-R scores in

Table 4. *Mean SDQ self-report scores with 95% confidence intervals by gender and country. Highest mean scores across countries indicated in boldface for boys and girls separately*

	Age		N^a	Emotional symptoms Mean (CI$_{95}$)	Conduct problems Mean (CI$_{95}$)	Hyperactivity and inattention Mean (CI$_{95}$)	Peer problems Mean (CI$_{95}$)	Prosocial behaviour Mean (CI$_{95}$)	Total difficulties Mean (CI$_{95}$)	Impact Mean (CI$_{95}$)
South Africa (present study)	12–16	Boys	1273	**3.26** (3.15–3.38)	**2.61** (2.48–2.73)	2.87 (2.71–3.03)	2.86 (2.69–3.03)	7.36 (7.20–7.51)	**11.59** (11.29–11.88)	0.70 (0.59–0.82)
		Girls	1965	**4.38** (4.23–4.54)	**2.28** (2.17–2.40)	2.64 (2.51–2.78)	**2.69** (2.53–2.85)	8.02 (7.90–8.15)	**12.00** (11.62–12.38)	0.71 (0.63–0.80)
		Total	3270	3.82 (3.72–3.93)	2.45 (2.34–2.55)	2.76 (2.63–2.88)	2.77 (2.62–2.93)	7.69 (7.58–7.80)	11.79 (11.52–12.07)	0.71 (0.63–0.78)
U.K.[b,c]	11–15	Boys	2135	2.6 (2.52–2.68)	2.4 (2.33–2.47)	**3.9** (3.85–3.95)	1.6 (1.54–1.66)	7.5 (7.43–7.57)	10.5 (10.28–10.72)	0.3 (0.29–0.31)
		Girls	2093	3.0 (2.91–3.09)	2.0 (1.93–2.07)	**3.6** (3.51–3.69)	1.4 (1.34–1.46)	**8.5** (8.44–8.56)	10.0 (9.77–10.23)	0.2 (0.19–0.21)
		Total	4228	2.8 (2.74–2.86)	2.2 (2.15–2.25)	3.8 (3.73–3.87)	1.5 (1.46–1.54)	8.0 (7.95–8.05)	10.3 (10.14–10.46)	0.2 (0.19–0.21)
Australia[b,c]	11–17	Boys	263	2.04 (1.81–2.27)	2.17 (1.95–2.39)	3.55 (3.27–3.83)	1.66 (1.54–1.78)	**7.58** (7.36–7.80)	9.37 (8.68–10.06)	
		Girls	290	2.75 (2.52–2.98)	1.50 (1.33–1.67)	2.85 (2.60–2.10)	1.40 (1.23–1.57)	**8.50** (8.33–8.67)	8.55 (7.92–9.18)	
		Total	553	2.41 (2.25–2.57)	1.82 (1.68–1.96)	3.18 (2.99–3.37)	1.52 (1.39–1.65)	8.06 (7.92–8.20)	8.94 (8.47–9.41)	
China[c,d]	11–17	Boys	326	2.36 (2.15–2.57)	2.35 (2.19–2.51)	3.65 (3.42–3.88)	**3.22** (3.04–3.40)	6.80 (6.59–7.01)	11.54 (11.00–12.08)	
		Girls	364	2.25 (2.05–2.45)	1.99 (1.85–2.13)	3.03 (2.82–3.24)	2.52 (2.36–2.68)	7.78 (7.60–7.96)	9.75 (9.28–10.22)	
		Total	690	2.30 (2.15–2.45)	2.16 (2.05–2.27)	3.32 (2.16–2.48)	2.85 (2.73–2.97)	7.32 (7.18–7.46)	10.60 (10.24–10.96)	

[a]For South Africa the number of observations reported is for total difficulties, but *n* may vary slightly across subscales due to missing data.
[b]http://www.sdqinfo.com/norms/UKSchoolNorm.html
[c]Confidence intervals (CI) estimated on the basis of standard deviations and number of observations with no adjustments for possible design effects.
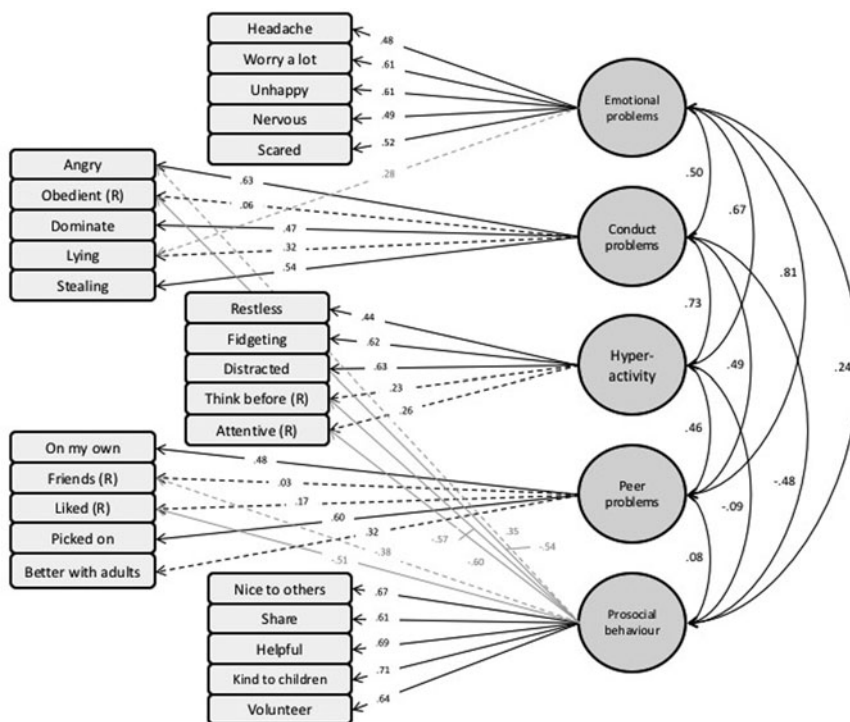[d]Du *et al.* (2008).

Fig. 1. Five-factor solution from confirmatory factor analysis (WLSMV estimator) adapted to data based on fit indices. $N = 3440$, number of clusters = 41. Dotted lines indicate significant loadings smaller than 0.40; $\chi^2 = 637.540$; d.f. = 258; $p = 0.000$; RMSEA = 0.021; CFI = 0.942; TIL = 0.933.

relation to a clinical benchmark (Goodman *et al.* 1998; van Roy *et al.* 2008). Given the limited clinical validation of the SDQ-S to date, we would recommend, at least as an interim strategy, that the newly-calculated South African cut-off values (based on the same 80–10–10 proportions used by the SDQ developers in the UK sample) be used for clinical and research purposes locally.

Internal consistency

Overall, we observed variable but fairly low internal consistency, with few gender differences in alpha values. Two Dutch studies also found that the internal consistency of the self-report version of the SDQ-S was low for the conduct and peer problems subscales (Muris *et al.* 2003; van Widenfelt *et al.* 2003). The subscale for peer problems has been reported to have low internal consistency in a number of studies (Goodman *et al.* 1998; Wolpert *et al.* 2015). There are a number of possible reasons for the relatively low alpha values observed here and in other studies. First, the small number of possible items to load onto the various subscales (maximum of 5 items) could explain the low internal consistency. Second, the clinical constructs of subdomains may represent inherently different psychological constructs, as in the

hyperactivity/inattention subdomain, for instance, where inattention, hyperactivity and impulsivity (that may all represent different neuropsychological constructs) come together for a diagnosis of ADHD. Last, as shown in other studies, removal of reversed items (where the response direction is opposite from other items), may have increased alpha values in our dataset (Di Riso *et al.* 2010; Essau *et al.* 2012). It would be important to explore the potential reason for the low internal consistency at an individual item level. For instance, low internal consistency has been thought to be attributable to the fact that some items that constitute the sub-scale may not fit the scale completely or that the sub-scales are only partly represented by the items (van Widenfelt *et al.* 2003).

Cross-country comparisons

The cross-country comparison showed that, in comparison to published UK, Australia and Chinese SDQ-S data, South African adolescent boys and girls had the highest mean scores on total difficulties as well as on the subscales of emotional symptoms and conduct problems. In contrast, South African boys and girls had the lowest mean scores for hyperactivity/inattention. The UK boys and girls had the highest mean scores for hyperactivity/inattention, while the

Australian sample had the highest scores for prosocial behaviours. The Chinese boys had the highest peer problem mean scores and Chinese boys and girls had the lowest means on prosocial behaviours. There may be many speculations about the potential reasons for these observed differences that may or may not find support in clinical practice. As raised above, some may argue that the bio-psychosocial context in Africa may explain the higher rates of mental health problems in the South African sample (Kashala *et al.* 2005). An alternative explanation comes from a more cultural and contextual model, suggesting that differences observed between countries could be explained by geographical location being a proxy for context and culture. To support this view, Ortuño-Sierra *et al.* (2015) reported similarities in SDQ-S scores among European countries that were in close proximity to one another, and similar in culture. However, there are also cross-country data that do not support this geographical proximity model (e.g. Essau *et al.* (2012)). The study by Stevanovic *et al.* (2014) arguably represented the most diverse cultural comparisons (India, Serbia, Nigeria, Turkey, Indonesia, Bulgaria and Croatia), and also showed variability in SDQ-S scores. It was interesting to observe that the Nigerian adolescents' emotional scores were also the highest, and hyperactivity/inattention the lowest in their study, but not all other scores showed a similar pattern that might suggest an 'African' pattern of adolescent strengths and difficulties. Apart from the psychometric recommendation that cross-country comparisons should be made with great caution (Goodman *et al.* 2012), and that in-country cut-offs for 'caseness' may be helpful, it remains a fascinating transcultural question whether culturally or contextually similar countries may have more similar psychometric profiles on mental health tools such as the SDQ.

SDQ-S factor structure in South Africa

Using South African data, we found limited support for the five-factor model. As shown in Fig. 1, only two sub-scales (emotional and prosocial) obtained significant loadings (higher than 0.40) for all five relevant items. The model as shown also contained a number of cross-loadings (loadings on 'wrong' indicators), suggesting that other dimensional structures may provide a better fit for South African data. High inter-correlations between some of the factors indicate that the number of factors can be reduced. Interestingly, emotional symptoms and prosocial behaviour were also the only two factors that had significant loadings for all five items in a study that assessed the SDQ teacher ratings in the Democratic Republic of Congo, another African country (Kashala *et al.* 2005). An

Australian study also examined the factor structure of the parent, teacher and self-report SDQ and found that the sub-scales of the SDQ were not unidimensional, suggesting that two or more factors were contributing to certain sub-scales in the instrument (Mellor & Stokes, 2007). Mellor & Stokes (2007) pointed out that the factor structure of the SDQ is conceptually made up of two independent factors – namely 'strengths' and 'difficulties'. This may suggest that prosocial behaviour and emotional symptoms may be the only two factors with significant loadings onto this strength (prosocial) *v.* difficulties (emotional) two-factor structure of the SDQ-S. We recommend further exploration of SDQ-S items from South Africa and other countries to see if different naturalistic behavioural clusters and factors may emerge in different socio-economic and cultural settings.

Limitations

We acknowledge several limitations in this study. First, we focused only on the SDQ-S. For overall clinical decision-making, self-reports should always be combined with information from other sources such as parental or teacher ratings, and future studies would benefit from inclusion of parent and teacher SDQs. However, the aim of this study was to examine the psychometric properties of this self-report version of the SDQ, given the potential usefulness of having direct access to the 'voices' of young people. We acknowledge that there may have been selection bias, given that participants had to have signed parental consent to participate. Nevertheless, the sample size was sufficient for the results presented in this publication, and was based on a random sample of public high schools across urban and rural parts of the Western Cape Province. In addition, we provided participants with the SDQ-S in three languages but did not ask them which language they used. Language may therefore have introduced biases in subtle ways. However, South Africa has 11 official languages and 'code-switching' (switching between different languages) is very common. For this reason we did not expect a significant language-based bias, but we acknowledge that this would benefit from empirical exploration. There is also the possibility that adolescents provided 'socially desirable' responses. However, questionnaires were completed anonymously, thus reducing the likelihood of social desirability as confounder.

Conclusion

Overall, our study suggested that the self-report SDQ may be a useful measurement instrument to identify adolescents at risk of mental health problems in a

South African setting. However, results showed that UK cut-off scores should not be used to determine 'caseness', and suggested that much further validation work is required to compare the normative cut-off values with gold standard clinical diagnostic instruments. Around the globe the SDQ-S may therefore be useful as an 'in-country' instrument, but results from our study and others suggest that it should be used with caution as a 'cross-country' comparative measurement tool. We propose that results to date suggest that cultural adaptation of the SDQ-S or subtle cultural, ethnic and pragmatic uses of language may be important predictors of over- or under identification of mental health difficulties across countries. Perhaps we need to return to a qualitative exploration of cultural use of mental health language in order to develop measurement instruments that capture similar global concepts appropriately and adequately in local settings.

Conflict of Interest

None.

Ethical Standards

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. In addition, 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.

Availability of Data and Materials

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 LE, Mathews C, Kaaya S, Katahoire AR, Onya H, Abraham C, Klepp K-I, Wubs A, Eggers SM, de Vries H (2014). Promoting sexual and reproductive health among adolescents in southern and eastern Africa (PREPARE): project design and conceptual framework. *BMC Public Health* **14**, 1–18.

Alem A, Kebede D, Fekadu A, Shibre T, Fekadu D, Beyero T, Medhin G, Negash A, Kullgren G (2009). Clinical course and outcome of schizophrenia in a predominantly

treatment-naive cohort in rural Ethiopia. *Schizophrenia Bulletin* **35**, 646–654.

Angold A, Costello EJ (2009). Nosology and measurement in child and adolescent psychiatry. *Journal of Child Psychology and Psychiatry* **50**, 9–15.

Bakare MO, Ubochi VN, Ebigbo PO, Orovwigho AO (2010). Problem and pro-social behaviour among Nigerian children with intellectual disability: the implication for developing policy for school based mental health problems. *Italian Journal of Pediatrics* **36**, 37.

Becker A, Rothenberger A, Sohn A, Ravens-Sieberer U, Klasen F, BELLA study group (2015). Six years ahead: a longitudinal analysis regarding course and predictive value of the Strengths and Difficulties Questionnaire (SDQ) in children and adolescents. *European Child and Adolescent Psychiatry* **24**, 715–725.

Bhola P, Sathyanarayanan V, Rekha DP, Daniel S, Thomas T (2016). Assessment of self-reported emotional and behavioral difficulties among pre-university college students in Bangalore, India. *Indian Journal of Community Medicine* **41**, 146–150.

Cortina MA, Fazel M, Hlungwani TM, Kahn K, Tollman S, Cortina-Borja M, Stein A (2013). Childhood psychological problems in school settings in rural Southern Africa. *PLoS ONE* **8**, e65041.

Deighton J, Croudace T, Fonagy P, Brown J, Patalay P, Wolpert M (2014). Measuring mental health and wellbeing outcomes for children and adolescents to inform practice and policy: a review of child self-report measures. *Child and Adolescent Psychiatry and Mental Health* **8**, 1–14.

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 Differences* **49**, 570–575.

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

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

Goodman R (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry* **38**, 581–586.

Goodman A, Goodman R (2011). Population mean scores predict child mental disorder rates: validating SDQ prevalence estimators in Britain. *Journal of Child Psychology and Psychiatry* **52**, 100–108.

Goodman R, Meltzer H, Bailey V (1998). The Strengths and Difficulties Questionnaire: a pilot study on the validity of the self-report version. *European Child and Adolescent Psychiatry* **7**, 125–130.

Goodman R, Ford T, Simmons H, Gatward R, Meltzer H (2000*a*). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *British Journal of Psychiatry* **177**, 534–539.

Goodman R, Renfrew D, Mullick M (2000*b*). Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *European Child and Adolescent Psychiatry* **9**, 129–134.

Goodman A, Heiervang E, Fleitlich-Bilyk B, Alyahri A, Patel V, Mullick MSI, Slobodskaya H, dos Santos DN, Goodman R (2012). Cross-national differences in questionnaires do not necessarily reflect comparable differences in disorder prevalences. *Social Psychiatry and Psychiatric Epidemiology* **47**, 1321–1331.

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

Kersten P, Czuba K, McPherson K, Dudley M, Elder H, Tauroa R, Vandal A (2016). A systematic review of evidence for the psychometric properties of the Strengths and Difficulties Questionnaire. *International Journal of Behavioural Development* **40**, 64–75.

Kieling C, Baker-Henningham H, Belfer M, Conti G, Ertem I, Omigbodun O, Rohde LA, Srinath S, Ulkuer N, Rahman A (2011). Child and adolescent mental health worldwide: evidence for action. *The Lancet* **378**, 1515–1525.

Kremer P, de Silva A, Cleary J, Santoro G, Weston K, Steele E, Nolan T, Waters E (2015). Normative data for the Strengths and Difficulties Questionaire for young children in Australia. *Journal of Paediatrics and Child Health* **51**, 970–975.

Mathews C, Eggers SM, Townsend L, Aarø LE, de Vries PJ, Mason-Jones AJ, De Koker P, McClinton-Appollis T, Mtshizana Y, Koech J, Wubs A, De Vries H (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 randomised controlled trial. *AIDS and Behavior* **20**, 1821–1840.

Mellor D (2004). Furthering the use of the Strengths and Difficulties Questionnaire: reliability with younger child respondents. *Psychological Assessment* **16**, 396–401.

Mellor D (2005). Normative data for the Strengths and Difficulties Questionnaire in Australia. *Australian Psychologist* **40**, 215–222.

Mellor D, Stokes M (2007). The factor structure of the strengths and difficulties Questionnaire. *European Journal of Psychological Assessment* **23**, 105–112.

Meltzer H, Gatward R, Goodman R, Ford T (2000). *Mental Health of Children and Adolescents in Great Britain*. The Stationery Office: London.

Menon A, Glazebrook C, Campain N, Ngoma MS (2007). Mental health and disclosure of HIV status in Zambian adolescents with HIV infection. *Epidemiology and Social Sciences* **46**, 349–354.

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

Myer L, Stein DJ, Jackson PB, Herman AA, Seedat S, Williams DR (2009). Impact of common mental health disorders during childhood and adolescence on secondary school completion. *South African Medical Journal* **99**, 354–356.

Needham B, Hill TD (2010). Do gender differences in mental health contribute to gender differences in physical health? *Social Science and Medicine* **71**, 1472–1479.

Ortuño-Sierra J, Fonseca-Pedrero E, Aritio-Solana R, Velasco AM, de Luis EC, 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, Heinz A, Walter H, Struve M, Gowland P, Paus T, Poustka L, Martinot J-L, Paillère-Martinot M-L, Vetter NC, Smolka MN, Lawrence C (2015). New evidence of factor structure and measurement invariance of the SDQ across five European nations. *European Child and Adolescent Psychiatry* **24**, 1523–1534.

Richter J, Sagatun Å, Heyerdahl S, Oppedal B, Røysamb E (2011). The Strengths and Difficulties Questionnaire (SDQ) – self-report. An analysis of its structure in a multiethnic urban adolescent sample. *Journal of Child Psychology and Psychiatry* **52**, 1002–1011.

Rønning JA, Handegaard BH, Sourander A, Mørch W (2004). The Strengths and Difficulties self-report questionnaire as a screening instrument in Norwegian community samples. *European Child and Adolescent Psychiatry* **13**, 73–82.

Skinner D, Sharp C, Marais L, Serekoane M, Lenka M (2014). Assessing the value of and contextual and cultural acceptability of the Strengths and Difficulties Questionnaire (SDQ) in evaluating mental health problems in HIV/AIDS affected children. *International Journal of Mental Health* **43**, 76–89.

Stevanovic D, Urbán R, Atilola O, Vostanis P, Singh Balhara YP, Avicenna M, Kandemir H, Knez R, Franic T, Petrov P (2014). 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* **30**, 1–12.

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.

van Widenfelt BM, Goedhart AW, Treffers PDA, Goodman R (2003). Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *European Child and Adolescent Psychiatry* **12**, 281–289.

Wolpert M, Cheng H, Deighton J (2015). Measurement Issues: review of four patient reported outcome measures: SDQ, RCADS, C/ORS and GBO – their strengths and limitation for clinical use and service evaluation. *Child and Adolescent Mental Health* **20**, 63–70.