International Journal of Applied Educational Research (IJAER)
Vol. 2 No. 2, 2023: 137 - 156

Factor Structure and Measurement Invariance of the Teacher Reported Strengths and Difficulties Questionnaires for Elementary School Students in the United States and China

Ruiqin Gao[1], Jin Liu[2], Ruyi Ding[3], Suzy Hardie[4], Wei He[5]
University of South Carolina
Corresponding Author: Ruiqin Gao rgao@email.sc.edu

A R T I C L E I N F O

Keywords: Factor Structure, Cross-Culture, Measurement Invariance, SDQ, Validity

Received : 17 Febuary
Revised : 19 March
Accepted: 29 April

A B S T R A C T

The Strengths and Difficulties Questionnaire (SDQ), a tool used by teachers, was examined in the current study for factor structure and measurement invariance in two culturally distinct contexts: China and the United States. Sixty-three elementary school pupils from the United States and China made up our sample. The three underlying factors of the SDQ—internalizing difficulties, externalizing problems, and prosocial behavior—were determined for each group sample based on the findings of confirmatory factor analysis. The SDQ demonstrated full configurational, metric, scalar, and residual invariance in elementary school environments in both China and the US, according to a multigroup confirmatory factor analysis. The study's findings added to the body of knowledge by confirming the reliability of teacher-reported SDQs and proving that it is feasible to use teacher-reported SDQs to compare the mental health of primary school kids from different cultural backgrounds.

INTRODUCTION

The Strengths and Difficulties Questionnaire (SDQ) is a widely used screening tool to assess behavioral characteristics in children aged 4 to 17 (Goodman, 2001). The Screening Depression Questionnaire (SDQ, n.d.) is a highly utilized tool available in over 80 languages and used in countries such as the Nordic countries, Germany, the Netherlands, China, and the United States.

Researchers looked into the psychometric qualities of the commonly used SDQ, such as measurement invariance (MI) and factor structure (Bibou-Nakou et al., 2019; Kersten et al., 2016). (Duinhof et al., 2020; Liang et al., 2019; Murray et al., 2021). Reliability of the factor structure of the SDQ was found to be inconsistent. To reevaluate the SDQ factor structure in different contexts, more research was needed. Historical research (Chiorri et al. (2016); Gomez and Stavropoulos (2020); He and others. 2013, Liang and colleagues. 2019; Murray and colleagues. , 2021; Rogge and colleagues. , 2018) investigated the measurement invariance between different groups (e.g. G. language versions, settings, time, parents, income, race/ethnicity, gender, age, and informants). As far as we are aware, no research has looked at the MI of the SDQ teacher version in elementary school settings in China or the United States. More scholars are attempting to comprehend the disparities in mental health between these two cultures as a result of the growing educational interaction between China and the United States. It is the responsibility of the researchers to confirm that educators in China and the United States understand the SDQ items similarly in this instance. Thus, the main goal of the current study was to investigate the factor structure and cultural MI of the SDQ teacher version in elementary school contexts in China and the United States.

LITERATURE REVIEW

Factor Structure of the SDQ

Goodman, 1997; Goodman et al., both used principal component analysis. , 1997; the original five-factor structure of the SDQ, which included the factors for emotional symptoms, conduct issues, peer problems, hyperactivity, and prosocial behavior, supported this original framework. as well as Essau et al. validates the factor analysis (CFA). 2012; both of them. 2013; Kersten along with associates. 2016, Liang and associates. 2019). Additionally, the three-factor structure was created. It was separated into three subscales: conduct problems, hyperactivity/inattention problems, and internalizing problems (emotional issues and peer relationship issues) (Bibou-Nakou et al. 2019; Riso with associates. 2010; Goodman along with associates. 2010).For high-risk samples, the five-factor structure should be used, and for community samples, the three-factor structure (Goodman and Goodman, 2009) should be used. Based on research involving Chinese participants, the self-reported SDQ has a 5-factor structure (Liang et al. 2019; Liu and colleagues. , 2013) and a 4-factor structure that includes the SDQ in parent and teacher versions (i. E. Problems with inattention, prosocial behavior, externalizing and internalizing issues, and Liu et al. in 2013). Research on the U. s. specimens revealed a 3-factor structure for the self-reported SDQ (Dickey & Blumberg, 2004) and a 5-factor structure (He et al. in 2013Most of the research to date has included children from early infancy

through puberty. The inconsistent outcomes showed how important it is to look at the SDQ factor structure using various samples. Most previous studies have employed self-or parent-reported SDQs. However, educators are more likely than parents to use a normative approach when assessing a child's behavior (Konold et al. 2004,). Furthermore, elementary school students may find it difficult to form reliable opinions about their behavior due to their limited cognitive abilities. The factor analysis of the teacher-reported SDQ for the SEB challenges of elementary school students has been limited by prior research.This study aimed to address these limitations by examining samples of elementary school students from China and the United States to explore the internal structure of the teacher-reported SDQ.

Measurement Invariance of the SDQ

According to Finney and Davis (2003), MI looks at whether group differences rather than variations in how the underlying term is conceptualized are the cause of variations in observed variables. Before comparing latent means between groups or time periods, MI analysis of latent constructs is necessary (Davidov, 2011). A common technique for assessing MI is confirmatory multigroup factor analysis (MG-CFA), which enables analysis of all measurement parameters (i.e. e. , factor structure, factor loadings, item intercept or threshold, and item residual variance) (Brown, 2015; Jung & Yoon, 2016).

Since SDQ is translated into multiple languages and adapted to different cultural contexts, it is important to optimize data equivalence across cultures. Cultural bias and non-comparable differences in mental health could be the cause of cross-national differences in SDQ scores (Goodman et al. 2012). However, there has been limited research on whether the latent structure that underpins the SDQ is consistent across national boundaries. Self-reported SDQ measurement invariance testing across European nations showed either no measurement invariance (Essau et al. either mild measurement invariance (Duinhof et al., 2012). 2020). Stevanovic along with others. (2015) was unable to locate a model that worked in countries in Europe, Asia, or AfricaMoreover, the MI of self-reported SDQ among European nations was the only MI examined in these cross-cultural studies. The mixed findings demonstrated the challenge of developing MI cross-culturally. We wanted to see if MI of teacher-reported SDQ can be established in different cultural contexts (e.g., the United States and China), as measures developed based on Western theories may be understood differently in Eastern culture. Furthermore, the prior investigations sampled exclusively adolescents. Because behavioral problems are predicted to decrease as children age (Campbell, 2002), SDQ should reveal different mental health issues in adolescents than in children.

The current study sought to address a research gap by examining the cultural MI of teacher-reported SDQ in Chinese and American elementary school settings. Two questions that the current study aims to address:

1) What is the latent component structure of teacher-reported SDQ in a sample of elementary school students from China and the United States?

2) Does the cultural MI of teacher-reported SDQ apply to China and the United States?

METHODOLOGY

Participants and Procedures

74 public elementary school teachers in China and 71 public elementary school teachers in the United States were asked to use the SDQ online Google Form from May to July 2021 in order to assess the social, emotional, and behavioral difficulties (SEB) of their students.. according to what they saw each day in the classroom, each teacher selected three students who may have been at a normal risk level, three at a medium risk level, and three at a higher SEB risk level, and then assessed these nine individuals using SDQ. Additional data were collected on children's gender, ethnicity and referrals to special services.

The sample included 1,305 instances of children aged 5 to 12 years old (M=9.2, SD=2.0), graded by 71 teachers from the United States (n=639) and 74 teachers from China (n=666). Table 1 contains demographic data for the children rated in the sample. Before data collection, Institutional Review Board permission and informed consent were acquired..

Table 1 *Demographic Characteristics of Participants(N=1,305)*

	U.S. Sample(N=639)	China Sample(N=666)
Variables	N (%)	N (%)
Gender		
Male	355(55.6%)	386(58.0%)
Female	283(44.4%)	280(42.0%)
Ethnicity		
Majority	276(43.2%)	659(98.9%)
Minority	363(56.8%)	7(1.1%)
Special Service		
No	475(74.3%)	588 (88.3%)
Yes	164(25.7%)	78(11.7%)
	Mean (SD)	Mean (SD)
Age	8.9(2.0)	9.5(1.9)

Instrument

The purpose of the 25-item SDQ is to assess the SEB challenges that kids and teenagers between the ages of 4 and 17 face. The anchor values on this three-point scale are "not true" (= 0), "somewhat true" (= 1), and "certainly true" (= 2). Five aspects of mental health issues are intended to be evaluated by the scale: emotional problems (e.g. G. "My headaches, stomachaches, or illnesses are frequent"), Behavior Issues (e.g. G. "I become very angry and frequently lose my temper"), issues related to hyperactivity and inattention (e.g. G. "I am fidgety and can't stay still for very long"), Peer Relationship Issues (e.g. G. "Alone is preferable to company of people my own age," and Prosocial Conduct (e.g. G. "I am gentle with younger kids.". Subscale scores are determined by summing the scores of pertinent items from each subscale—prosocial behavior excluded—after positively worded items from all subscales have been reverse coded. Subscale scores were added together without taking prosocial behavior into consideration, yielding a total difficulty score that ranged from zero to

forty. Greater levels of danger are indicated by higher difficulty scores. Based on risk estimates, instructors who took the SDQ had their cumulative scores split into three categories: normal (0–11), borderline (12–15), and abnormal (16–40). Prosocial Behavior scores that are higher are associated with lower risk. The English version (Goodman, 1997) and Chinese version of teacher-reported SDQ (https://www.sdqinfo.org/py/sdqinfo/b0.py) were used for the U.S. and Chinese samples, respectively.

Data Analysis

According to Muthén and Muthén (2019), Mplus 8.4 was used for all statistical analyses. The weighted least squares estimation approach with mean and variance adjustment (WLSMV) was employed in order to take into consideration the data's categorical nature (Finney & DiStefano, 2013). There were no missing data points discovered. To improve parameter estimations, Raykov and DiStefano (2021) suggested taking into account clustering effects caused by student nesting among teachers.

Confirmatory factor analysis

To find out if teacher ratings of Chinese and American elementary school students would reflect the previously determined factor structure, we used confirmatory factor analysis (CFA). We looked independently at the 5-factor, 4-factor, and 3-factor structures found in earlier studies for the Chinese and American populations because a shared measurement model can only be created if the suggested model fits both samples. We reverse-coded the Prosocial Behavior items in the same direction as those on the other four subscales to make the results more comprehensible. Higher scores denoted more prosocial issues.

The indices that are frequently used to analyze model fit include the p-value that is obtained using the WLSMV-based chi-square goodness-of-fit statistic, the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), the Root Mean sq\. Error of Approximation (RMSEA), and the Standardized Root Mean Linked sq\. (SRMR) (Finney and DiStefano, 2013). An appropriate model fit was indicated by the following cutoff values: RMSEA ≤.08, TLI ≥.90, CFI ≥.90, and SRMR ≤.10. A satisfactory model fit was indicated by the higher cutoff values, which were as follows: RMSEA ≤.05, SRMR ≤.08, TLI ≥.95, and CFI ≥.95 (Hu & Bentler, 1999). We also looked at the local fit indices because the size of the factor loadings and the number of items may have an impact on the global model fit (Greiff & Heene, 2017; McNeish et al., 2018). Factor loadings for the items should be more than 0.30, according to Costello and Osborne's (2005) recommendation. According to Raykov and Marcoulides (2012), standardized residuals larger than |3.0| can point to a local model mismatch. Model modification indices would be used to determine the incorrectly provided parameters if models did not demonstrate acceptable model fits. Researchers can find advice using model modification indices on a more complex model structure than what theory has predicted (MacCallum et al., 1992). Researchers can modify the initial hypothesized model to make it more parsimonious or better fitting by looking at parameter estimates, fit indices, and residuals. Once the update is complete, it is tested to see if it makes theoretical sense (Schreiber et al., 2006).

Reliability

We calculated Cronbach's alpha for the teacher-reported SDQ subscale scores for both groups separately (i.e., China and the U.STaking into account the average

covariance between item pairs, Cronbach's alpha computes the correlation between the scores on individual items and the overall scale score. Cronbach's alpha coefficients between 0.70 and 0.79 indicate acceptable internal consistency, values between 0.80 to 0.89 suggest good internal consistency, and values of 0.90 or above 0.90 indicate excellent internal consistency (Worm-Smeitink et al., 2017).

Convergent and discriminant validity

We looked at the teacher-reported SDQ's discriminant and convergent validity in the two cultures (China and the United States). The degree of correlation between several indicators of the same construct is measured by convergent validity (Ab Hamid et al. 2017). According to Hair et al., adequate convergent validity is indicated by an average variance extracted (AVE) value greater than 0 points 50. 2021). The degree to which constructs differ from one another empirically is evaluated by discriminant validity (Ab Hamid et al. 2017). As recommended by the Fornell-Lacker criteria method, we compared the correlation of latent variables with the square root of the AVE. Sufficient discriminant validity is indicated if the square root of each construct's AVE value is higher than the construct's correlations with other latent constructs (Hair et al. 2021).

Measurement invariance analysis

We evaluated configural, metric, scalar, and residual invariance as the four MI degrees successively using an MG-CFA model. Configurational invariance was examined by requiring that the factor structure in China and the USA match. After configural invariance was proven, the unstandardized factor loadings in the two nations were compared to assess metric invariance. We first checked for metric invariance and then looked at scalar invariance. To verify the residual variance's invariance, we contrasted two models. A model was fitted initially (i.e. e. , residual variance-invariance model A), where all of the Chinese group's residual variances were estimated freely. Next, we contrasted model A's residual variance invariance with model B, where the Chinese group's residual variances were all set to 1. Partial invariance can be confirmed by releasing the factor loading or intercept/threshold of one item at a time using modification indices if invariance cannot be demonstrated in any of the ensuing steps (Dimitrov, 2010).

The DIFFTEST tool in Mplus was used to estimate the chi-square change ($\Delta\chi2$) for statistical significance relative to the change in degrees of freedom for comparing nested models evaluate because WLSMV-based chi-square statistics vary by model and have standards. The $\Delta\chi2$ test has many drawbacks despite allowing for a statistical comparison of nest models. First, deviations from multivariate normality can have an impact on $\Delta\chi2$. Second, $\Delta\chi2$ is consistently statistically significant and substantial for large samples and/or complex models (Chen, 2007). Measurement invariance was evaluated using statistical model fit indices ($\Delta\chi2$) and the difference between the models' RMSEA values (Δ RMSEA), CFI values (ΔCFI), and SRMR values (ΔSRMR). The invariance hypothesis was deemed durable if the changes in these fit indices for the more constrained model satisfied the following requirements: for factor loading invariance, $\Delta CFI \leq 0.01$, $\Delta RMSEA \leq 0.015$, and $\Delta SRMR \leq 0.03$; and for item intercepts/thresholds and residual invariance for a sample size greater than 300: $\Delta CFI \leq 0.01$, $\Delta RMSEA \leq 0.015$ and $\Delta SRMR \leq 0.01$ and $\Delta CFI \leq 0.01$, $\Delta RMSEA \leq 0.015$ and $\Delta SRMR \leq 0.01$. We looked at modification indices in the noninvariance scenario to identify which factor loading(s) or thresholds for partial invariance should be freely estimated.

RESEARCH RESULT

Descriptive Statistics

Table 2 presents the proportions of response levels for each item for the U.S. sample and Chinese Sample. The results showed differences in the proportion of item response levels (e.g., Item 5, Item 11, Item 24) across two samples.

Table 2 *Descriptive Statistics of Teacher-reported SDQ Items*

	U.S. Sample			Chinese Sample		
Items	Item Levels			Item Levels		
	Not true	Somewhat true	Certainly true	Not true	Somewhat true	Certainly true
Item 1	13.3%	43.2%	43.5%	19.1%	47.0%	33.9%
Item 2	45.1%	31.3%	23.6%	43.7%	34.7%	21.6%
Item 3	64.3%	24.9%	10.8%	64.1%	25.5%	10.4%
Item 4	17.2%	38.5%	44.3%	14.1%	44.0%	41.9%
Item 5	64.8%	18.9%	16.3%	45.8%	34.5%	19.7%
Item 7	16.4%	35.8%	47.7%	22.4%	47.7%	29.9%
Item 8	57.3%	30.8%	11.9%	48.2%	42.2%	9.6%
Item 9	17.8%	32.6%	49.6%	10.2%	38.4%	51.4%
Item 10	45.9%	30.7%	23.5%	45.6%	31.2%	23.1%
Item 11	12.5%	29.3%	58.2%	8.0%	41.4%	50.6%
Item 12	64.3%	23.3%	12.4%	50.5%	35.0%	14.6%
Item 13	65.7%	24.4%	9.9%	53.5%	33.8%	12.8%
Item 14	11.0%	37.9%	51.2%	27.5%	42.2%	30.3%
Item 15	34.4%	33.2%	32.4%	34.2%	38.6%	27.2%
Item 16	62.4%	26.9%	10.6%	41.7%	44.6%	13.7%
Item 17	9.5%	35.1%	55.4%	10.4%	47.4%	42.2%
Item 18	67.1%	23.3%	9.5%	54.1%	35.1%	10.8%
Item 19	76.2%	18.8%	5.0%	63.8%	28.1%	8.1%
Item 20	17.5%	33.0%	49.5%	13.7%	42.2%	44.1%
Item 21	30.4%	42.3%	27.4%	19.1%	47.3%	33.6%
Item	86.4%	10.8%	2.8%	75.4%	19.2%	5.4%

22						
Item 23	54.1%	36.3%	9.5%	43.2%	42.5%	14.3%
Item 24	73.9%	17.8%	8.3%	53.9%	37.8%	8.3%
Item 25	32.4%	37.7%	29.9%	32.0%	40.7%	27.3%

Table 3 displays the findings of the CFA analysis. First, the five-factor solution was looked at. The U.S. sample's model suited the data quite well. However, because there was no convergence, the model did not perform well for the Chinese data. Next, we looked at the solution with four factors. The U.S. or Chinese sample data did not fit the answer. At last, we looked at the SDQ's three-factor structure. The model fit of the 3-factor models significantly improves in comparison to the 4-factor structure. However, Option 1: (Copy result)

At least one of the model fit indices was outside the advised bounds, indicating that Model 1 did not yet adequately fit either of the samples. According to modification indices, items 7, 11, 14, and 21 with loadings on various factors were implicated in model misfit.. In Model 2, we allowed Item 7 and Item 21 from Externalizing Problems, Item 11 and Item 14 from Internalizing Problems to load on Prosocial Behavior. Model 2 adequately fitted the U.S. sample. For the Chinese sample, the model fit index values are relatively near to the suggested bounds. Next, we looked up the local model fit data. We discovered that Item 23 ("Gets along better with adults than with other children") had loading values for both samples that were less than 0.30 (i.e., 0.22 and 0.00 for the Chinese and U.S. samples, respectively), indicating that this item did not accurately assess the constructs. As a result, we eliminated Item 23 from model 2 in model 3. Model 3 adequately fitted both samples. The resulting three-factor model is presented in Figure 1.

Table 3 *CFA Model Fit Statistics*

Model	**χ^2(df)**	**CFI**	**TLI**	**RMSEA (90% CI)**	**SRMR**
5 factors					
U.S. sample	1084.250*(265)	0.949	0.943	0.070(0.065-0.074)	0.098
China sample	No convergence				
4 factors					
U.S. sample	5159.074*(270)	0.697	0.664	0.168(0.164-0.172)	0.190
China sample	1607.006*(270)	0.803	0.781	0.086(0.082-0.090)	0.143
3 factors					
Model 1					
U.S. sample	1775.777*(272)	0.907	0.897	0.093(0.089-0.97)	0.138
China sample	1141.989*(272)	0.872	0.858	0.069(0.065-0.073)	0.178
Model 2					
U.S. sample	1055.840*(268)	0.951	0.945	0.068(0.064-0.072)	0.098
China	957.235*(268)	0.898	0.886	0.062(0.058-0.066)	0.105

sample					
Model 3					
U.S. sample	1005.179*(245)	0.953	0.948	0.070(0.065-0.074)	0.096
China sample	780.974*(245)	0.921	0.911	0.057(0.053-0.062)	0.093

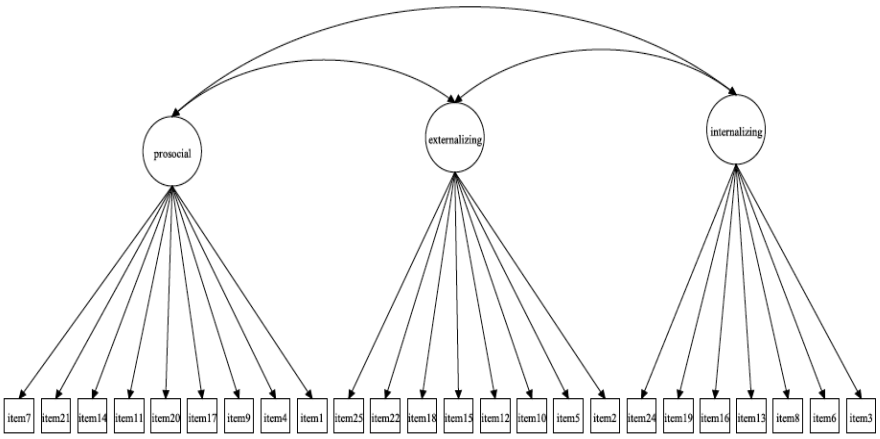
Note. χ2 = Chi-square test statistic, df = degree of freedom, CFI = comparative fit index, TLI = Tucker-Lewis Index, RMSEA = root mean squared error of approximation, CI = confidence interval, and SRMR = standardized root mean square residual.

Table 3 *CFA Model Fit Statistics*

Model	**χ^2(df)**	**CFI**	**TLI**	**RMSEA (90% CI)**	**SRMR**
5 factors					
U.S. sample	1084.250*(265)	0.949	0.943	0.070(0.065-0.074)	0.098
China sample	No convergence				
4 factors					
U.S. sample	5159.074*(270)	0.697	0.664	0.168(0.164-0.172)	0.190
China sample	1607.006*(270)	0.803	0.781	0.086(0.082-0.090)	0.143
3 factors					
Model 1					
U.S. sample	1775.777*(272)	0.907	0.897	0.093(0.089-0.97)	0.138
China sample	1141.989*(272)	0.872	0.858	0.069(0.065-0.073)	0.178
Model 2					
U.S. sample	1055.840*(268)	0.951	0.945	0.068(0.064-0.072)	0.098
China sample	957.235*(268)	0.898	0.886	0.062(0.058-0.066)	0.105
Model 3					
U.S. sample	1005.179*(245)	0.953	0.948	0.070(0.065-0.074)	0.096
China sample	780.974*(245)	0.921	0.911	0.057(0.053-0.062)	0.093

Note. χ2 = Chi-square test statistic, df = degree of freedom, CFI = comparative fit index, TLI = Tucker-Lewis Index, RMSEA = root mean squared error of approximation, CI = confidence interval, and SRMR = standardized root mean square residual.

Figure 1 *The Three-factor Structure of Teacher-Reported SDQ*



prosocial=Prosocial Behavior; externalizing=Externalizing Problems; Internalizing=Internalizing Problems.

Table 4 displays the significant relationships that were found between all items and the mentioned factors based on standardized factor loadings. There appears to be a positive linear relationship between the factors and the items because all loading values are positive.. Table 5 displays the range of correlations, from 0 points 18 to 0 points 54, between the three SDQ components that teachers reported. Internalizing problems and externalizing problems showed a somewhat positive correlation for all populations, indicating that children who are more prone to internalizing problems are also more likely to experience externalizing problems. There is more of a connection in the Chinese sample than in the US sample. The correlation coefficient between prosocial behavior and these issues, which ranges from 0.18 to 0.54, indicates that children who exhibit more positive prosocial behavior typically have fewer issues with internalizing or externalizing. The United States sample exhibits a stronger association than the Chinese sample.

Table 4 *3-factor CFA Results: Standardized Factor Loadings for Two Samples*

Factors and Items	U.S. Sample	Chinese Sample
Internalizing Problems		
3. Often complains of headaches, stomach pains or nausea	0.72	0.78
6. More of a loner, prefers to play alone.	0.60	0.81
8. Lots of worry or often seemed worried.	0.70	0.85
13. Frequently sad, depressed, or emotional..	1.00	0.89
16. Nervous or clingy in new situations, loses self-confidence easily.	0.70	0.84
19. Teased or bullied by other children.	0.76	0.82
24. Lots of fears, easy to scare.	0.68	0.76
Externalizing Problems		
2. Restless, overactive, cannot stay still for long.	0.80	0.82

5. Often loses temper.	0.84	0.81
10. Constantly fidgeting or squirming.	0.80	0.82
12. Often fights with other children or bullies them.	0.85	0.85
15. Easily distracted, concentration wanders.	0.85	0.85
18. Often lies or cheats.	0.84	0.83
22. Steals from home, school or elsewhere.	0.74	0.71
25. Good attention span, sees chores or homework through the end.	0.84	0.72
Prosocial Behavior		
1. Considerate of other people's feelings.	0.87	0.84
4. Shares readily with other children, for example toys, treats, pencils.	0.80	0.78
7. Generally well behaved, usually does what adults request.	0.89	0.62
9. Helpful if someone is hurt, upset or feeling ill.	0.83	0.78
11. Has at least one good friend.	0.74	0.75
14. Generally liked by other children.	0.85	0.87
17. Kind to younger children.	0.76	0.83
20. Often volunteers to help others (parents, teachers, other children).	0.81	0.82
21. Think things out before acting.	0.83	0.77

Note. All standardized factor loadings estimated are significantly above the $p<0.05$ level.

Reliability

For the Chinese group, the teacher-reported SDQ subscales (Internalizing Problems: α=0.89, Externalizing Problems: α=0.88, Prosocial Behavior: α=0.90) had Cronbach's alpha coefficients ranging from 0.88 to 0.90. The range of the United States' Cronbach's alpha coefficients. S. group ranged from 0 to 0 points (prosocial behavior: α=0.91, externalizing problems: α=0.87, and internalizing problems: α=0.82).Cronbach's alpha for the subscale scores exceeded 0.80, indicating good internal consistency. The results were highly similar between the two samples.

Convergent and Discriminant Validity

The average variance extracted (AVE) values for each construct from the U.S. are displayed in Table 5. S. and Chinese samples were above 0 point50, suggesting that the teacher-reported SDQ had sufficient convergent validity in both cultures. Additionally, Table 5 shows the correlation coefficients (off-diagonal) for each construct along with the square root of each AVE in the diagonal. For every one of the three constructs, the square root of the AVE values was higher than the correlations between the construct and other latent variables. The discriminant validity between the constructs when the teacher-reported SDQ was used in the U.S. is supported by this. S. and elementary school environments in China.

Table 5 *Convergent and Discriminant Validity Results*

	U.S. Sample		Chinese Sample	
Latent Variables	AVE	Latent Variables	AVE	Latent Variables

		A	B	C		A	B	C
Internalizing Problems (A)	0.56	**0.75**			0.68	**0.82**		
Externalizing Problems (B)	0.67	0.41	**0.82**		0.65	0.50	**0.80**	
Prosocial Behavior (C)	0.68	0.54	0.37	**0.82**	0.62	0.39	0.18	**0.79**

Note. The square root of the average variance extracted (AVE) (in bold) and correlations between constructs (off-diagonal).

Measurement Invariance Testing

Table 6 provides an overview of the results of the MI tests performed with MGCFA. Both the Chinese and American samples were subject to configural invariance, a three-factor structure discovered in the previous CFA analysis. The results showed that the internal structure of teacher-reported SDQ was the same in China and the United States (CFI=0.937, RMSEA=0.063, SRMR=0.095). The metric invariance model and the configuration model were compared. Regardless of the statistical significance of the scaled chi-square difference test (ΔCFI=0.004, ΔRMSEA=-0.003, ΔSRMR=0.006), the alternative fit measures demonstrated the feasibility of the DC hypothesis. To assess scalar invariance, we then contrasted this model with the metric invariance model. The chi-square showed a statistically significant shift. Using the same thresholds for each item, as shown by the alternative measures of fit modifications (ΔCFI=-0.003, ΔRMSEA=-0.001, ΔSRMR=0.003), the hypothesis could hold. Although there is a significant chi-square difference between the residual variance and invariance of Model A and Model B, the alternative fit measures showed that equal residual variances were maintained for the items (ΔCFI=0.001, ΔRMSEA=-0.002, ΔSRMR=0.006) , which supports the hypothesis of equal error variances for the items.

Table 6 *MGCFA Model Fit Across the U.S. and Chinese Samples(N=1,305)*

Model	χ^2 (df)	ΔSB-χ2(Δdf)	CFI	ΔCFI	RMSEA [90% CI]	ΔRMSEA	SRMR	ΔSRMR
Configural invariance	1763.845* (490)	____	0.937	___	0.063[0.060-0.066]	___	0.095	___
Metric invariance	1717.987* (511)	88.801* (21)	0.941	0.004	0.060[0.057-0.063]	-0.003	0.101	0.006
Scalar Invariance	1808.553(559)	156.370* (48)	0.938	-0.003	0.059[0.056-0.062]	-0.001	0.104	0.003
Residual Variance Invariance A	1876.737* (532)		0.934		0.062[0.059-0.065]		0.096	
Residual variance invariance B	1870.869* (556)	120.224* (24)	0.935	0.001	0.060(0.057-0.063)	-0.002	0.103	0.006

Note. MGCFA=multigroup confirmatory factor analysis; χ^2=Chi-square test statistic; df= degree of freedom; CFI=comparative fit index; RMSEA= root mean square error of approximation; CI= confidence interval; SRMR= standardized root mean square residual; Δ=difference. *p<.001.

The validity and reliability of an instrument ultimately determine its usefulness, so even though the SDQ is widely used in research, clinical, and community settings, it is important to continuously assess the psychometric qualities of the instrument. The SDQ's internal structure yields inconsistent results depending on the circumstance (e.g. g. , Nakou Bibou et coll. 2019; Kersten and associates. 2016, Liang et al. , 2019), even though earlier research looked at the SDQ's psychometric quality. Additionally, there isn't much experience in evaluating the SDQ's cross-cultural MI. Furthermore, while the majority of earlier studies focused on young children, the current study looked at elementary school students' teacher ratings. As untreated health issues can have a detrimental influence on parental relationships, physical health, and social participation, school-based general screening for mental health issues has been advised in order to identify children at risk early (Goodman et al. 2011; Pagliaccio and associates. 2012; Shonkoff and associates. 2012). Selecting a trustworthy screening tool is essential in this case. To support the validity evidence of the SDQ, we therefore looked at the factor structure and measurement invariance of the teacher-reported SDQ.

The basic three-factor structure of the scale was found in previous research (Bibou-Nakou et al. 2019; Riso and associates. ; Blumberg and Dickey (2004). 2010; Goodman & colleagues. 2010), which the CFA results corroborated. In both samples, the latent variables of prosocial behavior, externalizing problems, and internalizing problems in the CFA model yielded a good model fit. The latent construct of internalizing problems was not adequately explained by item 23 ("Gets along better with adults than with other children"). The reason for the failure could be because teachers don't have enough access to compare a child's interactions and communication skills with adults to those of other kids their same age. In the United States or China, we recommended eliminating this question when evaluating primary school kids based on teacher-reported SDQ. Besides, Item11("Has at least one good friend.") and Item 14 ("Generally liked by other children") from Internalizing Problems, Item 7("Generally well behaved, usually does what adults request"), and Item 21(" Think things out before acting") from Externalizing Problems loaded on Prosocial Behavior.

As individuals' prosocial behaviors aimed at helping or benefiting others, such as helping, sharing, and comforting (Eisenberg et al., 2006; Malti & Dys, 2018), these four items were all conceptually related to prosocial behaviors. Specifically, Carlo and Randall (2002) proposed that compliant behavior (i.e., enacted in response to others' requests) was one of the prosocial behaviors. Therefore, Item 7 theoretically assesses this type of prosocial behavior. Item 11("Has at least one good friend.") and Item 14 ("Generally liked by other children") are related to friendship. As the target of prosocial behavior was friends (Padilla-Walker et al., 2015) and friendship provided frequent opportunities to be prosocial (Padilla-Walker & Christensen, 2011), having friends or being liked by other children may function well in assessing the prosocial behavior of children. Besides, peer acceptance was an important aspect of pro-social behavior (Sebanc, 2003). Therefore, Item 11 and Item 14 can be used to assess elementary school children's prosocial behavior. Item 21 ("Think things out before acting") indicates concern for others, which is the key aspect of prosocial behavior which is motivated by concern for others (Witteck & Bekkers, 2015). Besides, the findings from the current study were also supported by the previous studies on the factor structure of SDQ. For example, Item 11, Item 14, and Item 7 also loaded on prosocial behavior in the study on the use of SDQ in Italy (Di Riso et al., 2010). Item 14 and Item 21 were allowed to cross-load with prosocial factors in the study on SDQ in

Denmark (Niclasen, et al., 2013). Item 11 and Item 14 also loaded on the Prosocial Behavior scale in previous studies in Flanders (Mieloo et al., 2014) and the Netherlands (Van Leeuwen, et al., 2006). The loading of these items on prosocial behavior may also be due to cultural differences in language translation and understanding (Toh et al., 2008).

The internal consistency of the subscales above 0.80 suggests that the teacher-reported SDQ has response stability when used in the United States and China. Furthermore, the convergent and discriminant validity of teacher-reported SDQs were also supported in both cultures. The results added to the literature on the reliability and validity of the teacher-reported SDQ when used in elementary school contexts in both the United States and China.

To examine the differences in measurement parameters (item threshold, residual variances, and factor loadings) between the Chinese and American samples, we conducted an MI test. The teacher-reported SDQ's configural, metric, scalar, and residual variance invariance (i.e. e. the MGCFA results supported equality of factor loadings, factor patterns, thresholds, and residual variances. The full invariance results imply that the teacher-reported SDQ can be used to compare mental health in elementary school settings in China and the United States. This study enhanced the field's comprehension of the psychometric properties of the teacher-reported SDQ by adding to the body of evidence proving its validity.

The current study contains a number of shortcomings in spite of its merits and contributions. First off, the data used in this study was limited to a sample of kids graded by instructors between the ages of 5 and 12. Children in clinical settings or those in later age groups might not be able to use the data. In order to determine whether the same findings hold true in other contexts, future research may look into the measurement invariance of SDQ. Second, owing to low factor loadings, we were able to identify a three-factor structure with the removal of one item. To ascertain whether comparable factor loadings may be found in the samples from other nations and age groups, replication is required. Third, the present study investigated the MI of the teacher-reported SDQ in China and the United States. The gender, years of teaching experience, and educational background of the teacher raters may also have an effect on their ratings. Subsequent studies could examine the MI among educators with respect to their cultural traits.

REFERENCES

Ab Hamid, M. R., Sami, W., & Sidek, M. M. (2017, September). Discriminant validity assessment: Use of Fornell & Larcker criterion versus HTMT criterion. In Journal of Physics: Conference Series (Vol. 890, No. 1, p. 012163). IOP Publishing.

Bibou-Nakou, I., Markos, A., Padeliadu, S., Chatzilampou, P., & Ververidou, S. (2019). Multi-informant evaluation of students' psychosocial status through SDQ in a national Greek sample. Children and Youth Services Review, 96, 47-54. https://doi.org/10.1016/j.childyouth.2018.11.022

Brown, T. A. (2015). Confirmatory factor analysis for applied research. Guilford publications.

Campbell, S. B. (2002). Behavior problems in preschool children: Clinical and developmental issues. New York: Guilford Press.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement Invariance. Structural equation modeling: a multidisciplinary journal, 14(3), 464-504. https://doi.org/10.1080/10705510701301834

Chiorri, C., Hall, J., Casely-Hayford, J., & Malmberg, L. E. (2016). Evaluating measurement invariance between parents using the Strengths and Difficulties Questionnaire (SDQ). Assessment, 23(1), 63-74. https://doi.org/10.1177/1073191114568301 Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Practical assessment, research, and evaluation, 10(1), 7.

Davidov, E. (2011). Nationalism and constructive patriotism: A longitudinal test of comparability in 22 countries with the ISSP. International Journal of Public Opinion

Research, 23(1), 88-103. https://doi.org/10.1093/ijpor/edq031

Dever, B. V., Dowdy, E., Raines, T. C., & Carnazzo, K. (2015). Stability and change of behavioral and emotional screening scores. Psychology in the Schools, 52(6), 618-629. https://doi.org/10.1002/pits.21825

Dickey, W. C., & Blumberg, S. J. (2004). Revisiting the factor structure of the strengths and difficulties questionnaire: United States, 2001. Journal of the American Academy of

Child & Adolescent Psychiatry, 43(9), 1159-1167. https://doi.org/10.1097/01.chi.0000132808.36708.a9

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. Measurement and Evaluation in Counseling and Development, 43(2), 121-
149. https://doi.org/10.1177/0748175610373459

Di Riso, D., Salcuni, S., Chessa, D., Raudino, A., Lis, A., & Altoè, G. (2010). The Strengths and Difficulties Questionnaire (SDQ). Early evidence of its reliability and validity in a community sample of Italian children. Personality and Individual Differences, 49(6), 570-575. https://doi.org/10.1016/j.paid.2010.05.005

Duinhof, E. L., Lek, K. M., De Looze, M. E., Cosma, A., Mazur, J., Gobina, I., ... & Stevens, G.

W. J. M. (2020). Revising the self-report strengths and difficulties questionnaire for cross-country comparisons of adolescent mental health problems: the SDQ-R. Epidemiology and psychiatricsciences,29. https://doi.org/10.1017/s2045796019000246

Eisenberg, N., Fabes, R. A., & Spinrad, T. L. (2006). Handbook of child psychology (Vol. 3, pp. 646-718). Hoboken, NJ, USA: John Wiley & Sons.

Essau, C. A., Olaya, B., Anastassiou-Hadjicharalambous, X., Pauli, G., Gilvarry, C., Bray, D., ... & Ollendick, T. H. (2012). Psychometric properties of the Strength and Difficulties

Questionnaire from five European countries. International Journal of Methods in Psychiatric Research, 21(3), 232-245. https://doi.org/10.1002/mpr.1364

Finney, S. J., & Davis, S. L. (2003). Examining the invariance of the achievement goal questionnaire across gender. In annual meeting of the American Educational Research Association, Chicago, IL.

Finney, S., & DiStefano, C. (2013). Dealing with nonnormality and categorical data in structural equation modeling. A second course in structural equation modeling. Greenwich, CT: Information Age.

Gao, X., Shi, W., Zhai, Y., He, L., & Shi, X. (2013). Results of the parent-rated strengths and difficulties questionnaire in 22,108 primary school students from 8 provinces of China. Shanghai Archives of Psychiatry, 25(6), 364.

Gomez, R., & Stavropoulos, V. (2020). Malaysian parent ratings of the strengths and difficulties questionnaire: factor structure and measurement invariance across language versions, gender, informants, and race. Assessment, 27(8), 1971-1984. https://doi.org/10.1177/1073191118787284

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. Journal of child psychology and psychiatry, 38(5), 581-586. https://doi.org/10.1111/j.1469-7610.1997.tb01545.x

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

Goodman, A., & Goodman, R. (2009). Strengths and difficulties questionnaire as a dimensional measure of child mental health. Journal of the American Academy of Child & Adolescent

Psychiatry, 48(4), 400-403. https://doi.org/10.1097/chi.0b013e3181985068

Goodman, A., Heiervang, E., Fleitlich-Bilyk, B., Alyahri, A., Patel, V., Mullick, M. S., ... &

Goodman, R. (2012). Cross-national differences in questionnaires do not necessarily reflect comparable differences in disorder prevalence. Social psychiatry and psychiatric epidemiology, 47(8), 1321-1331. https://doi.org/10.1007/s00127-011-0440-2

Goodman, A., Joyce, R., & Smith, J. P. (2011). The long shadow cast by childhood physical and mental problems on adult life. Proceedings of the National Academy of Sciences, 108(15), 6032-6037. https://doi.org/10.1073/pnas.1016970108

Goodman, A., Lamping, D. L., & Ploubidis, G. B. (2010). When to use broader internalising and externalising subscales instead of the hypothesised five

subscales on the Strengths and Difficulties Questionnaire (SDQ): data from British parents, teachers and children. Journal of abnormal child psychology, 38(8), 1179-1191.

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 & adolescent psychiatry, 7(3), 125-130. https://doi.org/10.1007/s10802-010-9434-x

Greiff, S., & Heene, M. (2017). Why psychological assessment needs to start worrying about model fit. European Journal of Psychological Assessment. 33(5), 313–317. https://doi.org/10.1027/1015-5759/a000450

Guo, M., Morawska, A., & Filus, A. (2018). Initial validation of the parent-report Child Adjustment and Parent Efficacy Scale (CAPES) in a Chinese cultural context. Assessment, 25(8), 1056-1073. https://doi.org/10.1177/1073191116681493

Hair Jr, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2021). A primer on partial least squares structural equation modeling (PLS-SEM). Sage publications.

He, J. P., Burstein, M., Schmitz, A., & Merikangas, K. R. (2013). The Strengths and Difficulties Questionnaire (SDQ): the factor structure and scale validation in US adolescents. Journal of abnormal child psychology, 41(4), 583-595. https://doi.org/10.1007/s10802-012-9696-6

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6(1), 1-55. https://doi.org/10.1080/10705519909540118

Jung, E., & Yoon, M. (2016). Comparisons of three empirical methods for partial factorial invariance: Forward, backward, and factor-ratio tests. Structural Equation Modeling: A Multidisciplinary Journal, 23(4), 567-584. https://doi.org/10.1080/10705511.2015.1138092

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 Behavioral Development, 40(1), 64-75. https://doi.org/10.1177/0165025415570647

Kettler, R. J., Glover, T. A., Albers, C. A., & Feeney-Kettler, K. A. (2014). An introduction to universal screening in educational settings. In R. J. Kettler, T. A. Glover, C. A. Albers, &

K. A. Feeney-Kettler (Eds.), Universal screening in educational settings: Evidence-based decision making for schools (pp. 3–16). Washington, DC: American Psychological Association.

Klasen, H., Woerner, W., Rothenberger, A., & Goodman, R. (2003). German version of the Strength and Difficulties Questionnaire (SDQ-German)--overview and evaluation of initial validation and normative results. Praxis der Kinderpsychologie und Kinderpsychiatrie, 52(7), 491-502.

Konold, T. R., Walthall, J. C., & Pianta, R. C. (2004). The behavior of child behavior ratings:
Measurement structure of the Child Behavior Checklist across time, informants, and child
gender. Behavioral Disorders, 29(4), 372-383.
https://doi.org/10.1177/019874290402900405
Liang, Lijuan, Juan Yang, and Shuqiao Yao. Measurement equivalence of the SDQ in Chinese adolescents: a horizontal and longitudinal perspective. Journal of affective disorders 257 (2019): 439-444. https://doi.org/10.1016/j.jad.2019.06.049
Liu, S. K., Chien, Y. L., Shang, C. Y., Lin, C. H., Liu, Y. C., & Gau, S. S. F. (2013). Psychometric properties of the Chinese version of Strength and Difficulties Questionnaire. Comprehensive Psychiatry, 54(6), 720-730. https://doi.org/10.1016/j.comppsych.2013.01.002
MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: the problem of capitalization on chance. Psychological bulletin, 111(3), 490. https://doi.org/10.1037/0033-2909.111.3.490
Malti, T., & Dys, S. P. (2018). From being nice to being kind: Development of prosocial behaviors. Current Opinion in Psychology, 20, 45-49. https://doi.org/10.1016/j.copsyc.2017.07.036
McNeish, D., An, J., & Hancock, G. R. (2018). The thorny relation between measurement quality and fit index cutoffs in latent variable models. Journal of personality assessment, 100(1), 43-52. https://doi.org/10.1080/00223891.2017.1281286
Mellor, D., Cheng, W., McCabe, M., Ling, M., Liu, Y., Zhao, Z., ... & Xu, Y. (2016). The use of the SDQ with Chinese adolescents in the clinical context. Psychiatry research, 246, 520-526. https://doi.org/10.1016/j.psychres.2016.10.034
Mieloo, C. L., Bevaart, F., Donker, M. C., van Oort, F. V., Raat, H., & Jansen, W. (2014). Validation of the SDQ in a multi-ethnic population of young children. The European Journal of Public Health, 24(1), 26-32. https://doi.org/10.1093/eurpub/ckt100
Murray, A. L., Speyer, L. G., Hall, H. A., Valdebenito, S., & Hughes, C. (2021). Teacher versus parent informant measurement invariance of the Strengths and Difficulties
Questionnaire. Journal of pediatric psychology, 46(10), 1249-1257.
https://doi.org/10.31234/osf.io/u89rb Muthén, L. K., & Muthén, B. (2019). Mplus. The comprehensive modelling program for applied researchers: user's guide, 5.
Niclasen, J., Skovgaard, A. M., Andersen, A. M. N., Sømhovd, M. J., & Obel, C. (2013). A confirmatory approach to examining the factor structure of the Strengths and Difficulties Questionnaire (SDQ): a large scale cohort study. Journal of abnormal child psychology, 41, 355-365. https://doi.org/10.1007/s10802-012-9683-y

Obel, C., Heiervang, E., Rodriguez, A., Heyerdahl, S., Smedje, H., Sourander, A., ... & Olsen, J. (2004). The strengths and difficulties questionnaire in the Nordic countries. European child & adolescent psychiatry, 13(2), ii32-ii39.

Padilla-Walker, L. M., & Christensen, K. J. (2011). Empathy and self-regulation as mediators between parenting and adolescents' prosocial behavior toward strangers, friends, and family. Journal of Research on Adolescence, 21(3), 545-551. https://doi.org/10.1111/j.1532-7795.2010.00695.x

Pagliaccio, D., Luby, J., Gaffrey, M., Belden, A., Botteron, K., Gotlib, I. H., & Barch, D. M. (2012). Anomalous functional brain activation following negative mood induction in children with pre-school onset major depression. Developmental Cognitive Neuroscience, 2(2), 256-267. https://doi.org/10.1016/j.dcn.2011.11.008

Raykov, T., & DiStefano, C. (2021). Design Effect in Multilevel Settings: A Commentary on a Latent Variable Modeling Procedure for Its Evaluation. Educational and Psychological Measurement, 00131644211019447. https://doi.org/10.1177/00131644211019447

Raykov, T., & Marcoulides, G. A. (2012). A first course in structural equation modeling. Routledge.

Rogge, J., Koglin, U., & Petermann, F. (2018). Do they rate in the same way? Testing of measurement invariance across parent and teacher SDQ ratings. European Journal of

Psychological Assessment, 34(2), 69. https://doi.org/10.1027/1015-5759/a000445

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. The Journal of educational research, 99(6), 323-338. https://doi.org/10.3200/joer.99.6.323-338

SDQ: Information for researchers and professionals about the strengths & difficulties questionnaires. n.d., https://www.sdqin fo.com/.

Sebanc, A. M. (2003). The friendship features of preschool children: Links with prosocial behavior and aggression. Social development, 12(2), 249-268. https://doi.org/10.1111/1467-9507.00232

Shonkoff, J. P., Garner, A. S., Siegel, B. S., Dobbins, M. I., Earls, M. F., McGuinn, L., ... &

Committee on Early Childhood, Adoption, and Dependent Care. (2012). The lifelong effects of early childhood adversity and toxic stress. Pediatrics, 129(1), e232-e246. https://doi.org/10.1542/9781610023658-part02-the_lifelong

Stevanovic, D., Urbán, R., Atilola, O., Vostanis, P., Balhara, Y. S., Avicenna, M., ... & Petrov, P. (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(4), 323-334. https://doi.org/10.1017/s2045796014000201

Toh, T. H., Chow, S. J., Ting, T. H., & Sewell, J. (2008). Chinese translation of strengths and difficulties questionnaire requires urgent review before field

trials for validity and reliability. Child and Adolescent Psychiatry and Mental Health, 2(1), 1-4. https://doi.org/10.1186/1753-2000-2-23

Van Leeuwen, K., Meerschaert, T., Bosmans, G., De Medts, L., & Braet, C. (2006). The

Strengths and Difficulties Questionnaire in a community sample of young children in Flanders. European Journal of Psychological Assessment, 22(3), 189-197.https://doi.org/10.1027/1015-5759.22.3.189

Van Widenfelt, B. M., Goedhart, A. W., Treffers, P. D., & Goodman, R. (2003). Dutch version of the Strengths and Difficulties Questionnaire (SDQ). European child & adolescent psychiatry, 12(6), 281-289. https://doi.org/10.1007/s00787-003-0341-3

Wittek, R., & Bekkers, R. H. F. P. (2015). The sociology of prosocial behavior and altruism. In International Encyclopedia of Social and Behavioral Sciences (pp. 579-583). Elsevier.https://doi.org/10.1016/b978-0-08-097086-8.32158-4

Worm-Smeitink, M., Gielissen, M., Bloot, L., Van Laarhoven, H. W. M., Van Engelen, B. G.

M., Van Riel, P., ... & Knoop, H. (2017). The assessment of fatigue: Psychometric qualities and norms for the Checklist individual strength. Journal of psychosomatic research, 98, 40-46. https://doi.org/10.1016/j.jpsychores.2017.05.007