

DOI: 10.5455/msm.2018.30.164-169

Received: June 29 2018; Accepted: August 23, 2018

© 2018 Brikena Allkoja

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ORIGINAL PAPER

Mater Sociomed. 2018 Aug; 30(3): 164-169

# Psychometric Properties of the Strengths and Difficulties Questionnaire in a Sample of Albanian Elementary School Children

Brikena Allkoja

Department of  
Neuroscience, University  
Hospital Center "Mother  
Teresa" Tirana, Albania

**Corresponding author:**  
Brikena Allkoja, UHC  
"Mother Tereza",  
Tirana, Albania. <https://orcid.org/0000-0001-9678-1646>. Mobile:  
+(355)672074269.  
E-mail: [brikenaallkoja@yahoo.com](mailto:brikenaallkoja@yahoo.com).

## ABSTRACT

**Introduction:** The Strengths and Difficulties Questionnaire (SDQ) is a 25-item screening measure for emotional and behavioral problems in children and adolescents ages 4 to 17. **Aim:** To evaluate the reliability and factor structure of SDQ teacher version in a sample of Albanian elementary school children. **Material and Methods:** Teachers of children (N = 542) from three elementary schools of Tirana city, Albania, (aged 6 to 12 years, grade 1-4) completed the Albanian version of SDQ. Internal consistency reliability was determined by calculation of the Cronbach- $\alpha$  coefficient and average inter-item correlation. Both Confirmatory factor analysis (CFA) and Exploratory factor analysis (EFA) examined the underlying factor structure of the questionnaire. **Results:** Internal consistency reliability of subscales ranged from "acceptable" to "good" with  $\alpha$  values that ranged from  $\alpha = .68$  (prosocial) to  $\alpha = .83$  (hyperactivity-inattention) and average inter-item correlation ranged from 0.5 for hyperactivity scale to 0.20 for total difficulties scale. Confirmatory factor analysis did not confirm the five factor structure of the teacher version of SDQ in Albanian, with values of incremental fit indices ( $\chi^2(265) = 1562.72, p < .0001$ ; Comparative Fit index, (CFI) = .814; Tucker-Lewis Index, (TLI) = .772) and Root Mean-Square Error of Approximation (RMSEA) = .095 indicating a misfit. Exploratory factor analysis revealed a different pattern of factor loadings for items of hyperactivity and conduct scale suggesting a non-one-dimensional contribution of the items. **Conclusion:** The SDQ demonstrated satisfying reliability based on internal consistency coefficients but must be used with caution when screening for Attention Deficit/Hyperactivity disorder in Albanian elementary school children.

**Keywords:** children, primary school, psychometric, Strengths and Difficulties, Questionnaire.

## 1. INTRODUCTION

Epidemiological data indicate that 20-40% of all children worldwide have a clinically diagnosable disorder, which often manifests as internalizing or externalizing behaviors (1-3) and that many more children exhibit specific symptoms or subclinical problems (4-6).

Despite the prevalence of mental health disorders in youth, only 45% of youth with a mental health diagnosis receive treatment, and only 24% of those individuals receive care in the school system (7). Within the school system, universal screening for mental health difficulties has been recommended as a way to identify children who are experiencing difficulties and provide the services necessary to support them (8). The goal of universal screening is to identify childhood problems before the behaviors exceed the threshold for a parent or teacher referral for services (9). School-based screening is advantageous because of the natural aggregation of most children at the transition to primary school and the accessibility of both teachers and parents as informants which may increase the accuracy of identification of children with problems (10). Within the framework of implementing a program of school-based universal screening for child psychopathology, it is of paramount importance the selection of an appropriate universal screening tool (11). When selecting a screening tool, planning teams must consider the domain of interest, psychometric adequacy, feasibility, acceptability and overall utility of the screening tool (12). In this context, considering the Strengths and Difficulties Questionnaire (SDQ) (13) as a broad-band valid screening tool for childhood mental health difficulties is extensively supported by research studies (14). The Strengths and Difficulties Questionnaire is a brief behavioral screening questionnaire that

can be completed by parents and teachers of 4-17 years old and youths 11-16 years old. SDQ consists of 25 items rated on a Likert-type scale and equally divided into five subscales: emotional symptoms, conduct problems, hyperactivity-inattention, peer problems, and prosocial behavior. The sum of the four subscales generates a total difficulties score. SDQ has several features that make it a useful screening measure for child mental health problems. It is freely available, takes only some minutes to be completed and it is translated in 85 languages, including Albanian. Additionally, since parent and teacher versions are identical, they enable the collection of the information about specific behaviors from different informants. Finally, the SDQ covers a broad range of common childhood difficulties, including internalizing, externalizing, inattention/hyperactivity, and peer relationship problems so that information on several aspects of children's functioning can be gathered with a single, brief measure (14). SDQ parent, teacher and/or self-report versions provide a five-factor structure, *i.e.*, emotional symptoms, conduct problems, hyperactivity/inattention, peer problems and prosocial behavior scale, that corresponds with the domains of psychopathology and personal strengths it intends to measure. Psychometric Properties of the Parent and Teacher SDQ has been investigated in several studies across different countries worldwide (14-18, 27) with mixed results. The reliability of the SDQ can be defined as satisfactory, although it should be noted that the internal consistency of the teacher ratings might be higher except for the peer problem scale having a value below 0.70 (14). However, up to date, no study has investigated the psychometric properties of SDQ in a population of Albanian youths. The present study is the first one to investigate the psychometric properties of the teacher-reported SDQ in an Albanian population of elementary school children.

## 2. AIM

To evaluate the reliability and factor structure of SDQ teacher version in a sample of Albanian elementary school children.

## 3. MATERIAL AND METHODS

### Participants and Procedure

Permission to conduct a research in the public primary schools was first applied for, to the Regional Directory of Education which further granted permission to conduct the study on three public primary schools of one district of Tirana city. After some preliminary meetings with the Principals and teachers of the selected schools, the consenting teachers and students were provided with separate envelopes for parents and teachers, containing a cover letter, informed consent and questionnaires to be completed. Inclusion criteria were children grade 1-4, with no hearing, seeing or major physical impairment. Participants were parents and teachers of the children (N =893) from grade 1 to 4 of the three schools, aged 6 to 11 (M=8.85; SD=1.2). Parents returned signed informed consent and completed questionnaires for 546 (61%) children, while teachers returned completed questionnaires for 542 (60%) children. 451 (50.5%) children had teachers and parents completed questionnaire. Overall, there were returned completed

questionnaire from parents or teachers for 637 (71%) children (girls =320, boys= 317). Of 32 teachers, only 24 (75%) consented to participate in the study.

### Measures

Teacher official version of the Albanian translated SDQ (www.sdq.org) was used as a measure tool and teachers were asked to think about the child behavior during the last six months and to select the alternative answer that best described the child. They were encouraged to answer all the items even if they were not sure about the best alternative. Ratings are scored as follows: not true =0, some-what true =1, and certainly true =2. Five items are reversely scored (certainly true =2, somewhat true =1, not true =1). Scoring for each subscale was calculated by summing ratings for each of their respective items. Total score was calculated summing the scores of the four problem subscales (emotional symptoms, conduct problems, hyperactivity/impulsivity and peer problems).

### Statistical analysis

The reliability of the subscales for teacher SDQ ratings was assessed using Cronbach-alpha coefficients and the average inter-item correlation which examines the extent to which scores on one item are related to scores on all other items in a scale. Ideally the Cronbach alpha should be greater than 0.7 (19) with  $\alpha = 0.70$  and below considered low, values between  $\alpha = 0.70 - \alpha = 0.80$  considered acceptable and values  $\alpha = 0.80$  and above good (20). SDQ subscales contain five items each and since it is not uncommon with short scales with less than ten items to find low Cronbach alpha values, it is reasonable and more appropriate to use an alternative measure of internal consistency such as the averages inter - correlations of items within a scale (19). Briggs and Cheek (21), recommend that the optimum range of the average inter-item correlation for a set of items should be between .20 and .40. We further explored the underlying structure of the teacher SDQ version using both confirmatory factor analysis (CFA) and exploratory factor analysis (EFA). To evaluate model fit, CFA was conducted, and a range of fit statistics were calculated. For the incremental fit statistics the Tucker-Lewis Index (TLI) and the Comparative Fit index (CFI) values less than .90 indicate lack of fit, values between .90 and .95 indicate reasonable fit and values between .95 and 1.00 indicate good fit (22). Byrne (23) describes the Root Mean-Square Error of Approximation (RMSEA) as the most informative statistic in determining model fit as it takes into account the number of variables that are estimated in the model. RMSEA values .05 or lower indicate good fit, values between .05 and .08 indicate reasonable fit. Factor loading values of .32 or more were required for the variable to be interpreted; values  $> .45$  were considered fair;  $> .55$  were considered good;  $> .63$  were considered very good and  $> .71$  were considered excellent (24). After the sample was first assessed for its suitability for factor analysis with Bartlett's Test of Sphericity and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy, to further investigate the number of constructs and structure of this measure, an EFA using Principal Component Analysis (PCA) used to extract the factors followed by oblique rotation of factors using Oblimin rotation ( $\delta = 0$ ) was conducted using SPSS version 24. To establish the

number of factors to be retained we first applied Kaiser’s criterion (eigenvalues above 1), along with the inspection of the screeplot, and the use of Horn’s parallel analysis (25) using the software Monte Carlo PA, developed by Watkins (26). Only factors with eigenvalues exceeding the values obtained from the corresponding random data set were retained for further investigation.

**4. RESULTS**

**Reliability**

Cronbach’s alpha values (Table 1) for teacher-rated SDQ subscales ranged from .67 (prosocial problems) to .83 (hyperactivity-inattention) while average inter-item correlations fell within the recommended optimum range and ranged from 0.20 for total difficulties scale to 0.50 for hyperactivity scale indicating acceptable to good internal consistency.

**Confirmatory factor analysis**

We used AMOS 24 software to perform a CFA based on the data we obtained from teachers for a total of 542 children. We chose maximum likelihood estimation because our data were normally distributed. The theoretical five factor model proposed by authors is presented in Figure 1. The chi-square test was significant ( $\chi^2(265)=1552.549$ ;  $p<0.001$ ) and the other fit indices (CFI= .814, TLI= 0.772; RMSEA= .095) indicated a misfit of the data to the five-factor model.

Unstandardized estimates together with standardized parameter estimates, and the squared multiple correlation (SMC) values are shown in Table 2. The highest factor loading was found for the item on the prosocial scale (helpful = 0.99) and the lowest but still in the interpretable range was on the conduct scale (obedient=0.39). The squared multiple correlation (SMC) values indicate (lower bound) the reliability of the measure; “fights” (.92) and “obedient” (.15) have the highest and lowest, respectively. A range of alternative factor solutions were tried and no support was found for the three factor solution proposed from Dickey and Blumberg (2004), while other alternative modifications resulted in poor model fits.

**Exploratory factor analysis**

The samples were first assessed for their suitability for factor analysis. Bartlett’s Test of Sphericity was highly significant ( $p < .001$ ) and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (KMO=0.810) value supported the factorability of the matrix (21). Principal components analysis revealed six eigenvalues exceeding 1, explaining 57.38 % of the variance. Despite the initial extraction resulted in six components with eigenvalues greater than 1, only the first five components exceeded the criterion value obtained from Parallel Analysis (26) and this was further supported from the inspection of the screeplot. So the num-

	Emotional problems	Conduct problems	Hyperactivity-inattention	Peer problems	Prosocial behaviour	Total difficulties
Cronbach- $\alpha$	0.79	0.71	0.83	0.76	0.67	0.83
Mean Inter-item correlation	0.45	0.37	0.50	0.4	0.30	0.20

Table 1. Internal Consistency of SDQ subscales

Item	Scale	Loadings (unstandardized)	Standard error	p	Loadings (standardized)	Squared Multiple Correlations
24. Easily Scared	Emotional	1.00			.50	.25
16. Nervous	Emotional	1.36	.13	***	.66	.44
13. Unhappy	Emotional	1.24	.12	***	.76	.58
8. Worries	Emotional	1.35	.12	***	.81	.65
3. Somatic	Emotional	1.09	.11	***	.63	.39
22. Steals	Conduct	1.000			.44	.19
18. Lies Or Cheats	Conduct	1.54	.21	***	.40	.16
12. Fights	Conduct	4.48	.43	***	.96	.92
7. Obedient	Conduct	2.22	.31	***	.40	.15
5. Tempers	Conduct	3.86	.38	***	.80	.64
25. Good Attention	Hyper	1.000			.90	.80
21. Thinks Before Acting	Hyper	.77	.05	***	.69	.48
15. Distracted	Hyper	.98	.04	***	.82	.68
10. Fidgets	Hyper	.57	.05	***	.52	.27
2. Restless	Hyper	.62	.06	***	.50	.25
23. Better With Adults	Peer	1.00			.40	.17
19. Picked On Or Bullied	Peer	.76	.10	***	.52	.27
14. Others Like	Peer	2.05	.24	***	.82	.67
11. Good Friend	Peer	1.85	.22	***	.77	.59
6. Alone	Peer	1.42	.18	***	.61	.37
20. Volunteers	Posocial	1.00			1.00	.99
17. Kind To Child	Posocial	.46	.04	***	.45	.20
9. Helpful	Posocial	.98	.01	***	.98	.97
4. Shares Readily	Posocial	.55	.05	***	.47	.22
1. Considerate	Posocial	.42	.04	***	.39	.15

Table 2. Unstandardized and standardized loadings for CFA

ber of factors that we retained for further analysis was five.

The PCA with oblimin rotation of teacher ratings (Table 3) revealed a different five - component pattern matrix.

The items of the three SDQ subscales of emotional problems, peer problems and prosocial behavior subscales loaded in the predicted subscales and the loadings fell within the “good” to “excellent” range. Items of two other subscales, specifically items of conduct problems and hyperactivity/inattention scales loaded differently from the predicted SDQ model, with some cross-loadings in more than one component. So, the items, 2 (“restless”), 10 (“fidgets”), 5 (“tempers”), of the hyperactivity scale and items 12 (“fights”) and 7 (“obedient”), loaded strongly within the same component, with the exception of “obedient”, which loaded poorly in this component and cross loaded also poorly on the component of prosocial scale. Furthermore, the rest of the items from hyperactivity scale, 15 (“distracted”), 25 (“good attention”), 21 (“thinks before acting”) and conduct scale 18 (“lies or cheats”), loaded in one separate component and item 22 (“steals”) loaded fairly (item 22=.44) in the component of emotional scale. Examination of structure matrix did not reveal a clear five factor solution.

Items	Pattern Matrix					Structure Matrix					Co.a
	1	2	3	4	5	1	2	3	4	5	
8. Worries	0.80					0.80					0.666
13. Unhappy	0.76					0.77					0.608
3. Somatic	0.74					0.72					0.527
24. Easily Scared	0.69					0.69			0.42		0.570
16. Nervous	0.60					0.66				0.40	0.529
22. Steals	0.44					0.51					0.288
14. Others Like*		0.83					0.82				0.682
11. Good Friend*		0.79					0.79				0.636
6. Alone		0.73					0.74				0.556
19. Picked On Or Bullied		0.65					0.66				0.461
23. Better With Adults		0.57					0.56				0.328
20. Volunteers			0.80					0.84		-0.47	0.824
9. Helpful			0.79					0.84		-0.47	0.813
17. Kind To Child			0.69					0.68			0.484
4. Shares Readily			0.67					0.65			0.449
1. Considerate			0.61					0.65			0.473
2. Restless				0.80					0.83		0.724
10. Fidgets				0.80					0.82		0.748
5. Tempers				0.67		0.46			0.74		0.619
12. Fights				0.61		0.55			0.74		0.683
7. Obedient*			-0.40	0.40				-0.45	0.47		0.373
15. Distracted					0.83					0.86	0.758
25. Good Attention*					0.76					0.82	0.745
21. Thinks Before Acting*					0.63				0.41	0.70	0.603
18. Lies Or Cheats					0.42	0.46				0.52	0.397

<sup>a</sup> Communalities, \* Reversely scored items  
 Note: Major loadings (>0.40) for each item are presented

Table 3. Pattern And Structure Matrix for PCA with Oblimin Rotation of Five Factor Solution of SDQ teacher ratings

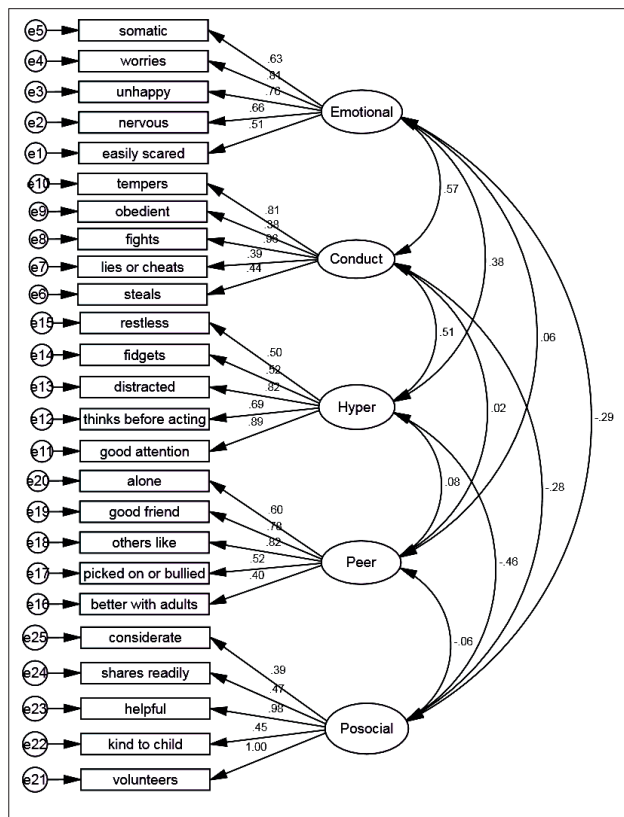


Figure 1 CFA path diagram of standardized estimates (hypothesized five – factor model)

### 5. DISCUSSION

This study examined the internal consistency of the Albanian SDQ version for teachers and both confirmatory and exploratory and factor analysis were performed to examine the fit of the obtained data to the proposed five factor model. Results from a number of studies show acceptable internal consistency for the SDQ subscales for teacher ratings (14) with an adequate internal consistency for the hyperactivity scale. In this study, Cronbach alpha values overall fell within the “acceptable to good range” with the exception of prosocial behavior scales that were considered low. The hyperactivity subscale had the highest value ( $\alpha = 0.83$ ) indicating an adequate fit. Average inter-item correlations values fell within the optimum range recommended for adequate internal consistency and hyperactivity scale had the highest value of 0.50. In conducting CFA, the five-factor model misrepresented the data for teacher ratings and despite attempted modifications no alternative structural model that could fit the data, could be detected. Exploratory factor analysis sustained a five-factor structure but revealed a different pattern of factor loadings for the items of conduct and hyperactivity scales, suggesting a shared contribution of the items in both factors and further a poor differentiation between these two factors. Other studies that have examined the factor structure of the SDQ for parent and teacher ratings have also failed to confirm the five-factor structure of this measure (14) suggesting that the instrument may be measuring different constructs across the groups (17). Hill and Hughes (27) found only marginal

baseline fit for a five-factor structure for either informant in their sample of US children (mean age six). Another study (16) also found a different pattern of loadings for the items of conduct scale and a tendency of positively worded items (“obedient”) to load on the prosocial scale. Additionally, other authors that have conducted researches in the field of child psychopathology have reported difficulties in the differentiating between hyperactivity and aggression. In an attempt to disentangle the potentially confounded domains of hyperactivity and conduct problems, Milich and Loney (28) began to derive independent dimensions of hyperactivity and aggression, resulting in a new measuring instrument designed to differentiate between the two dimensions of behavior, but surprisingly, the correlation between the two factors was  $r = .63$ , suggesting for an inherent association between these two domains (29). In another study, Schachar, Sandberg, Rutter (30), examined the agreement between teachers’ ratings and observations of hyperactivity, inattentiveness, and defiance, and found that defiance was more reliably rated than hyperactivity or inattentiveness. In particular, hyperactivity ratings may be influenced by observed defiance and inattentiveness, inattentiveness ratings by observed defiance, and behavior disorder by observed hyperactivity and inattentiveness. So, defiance toward a teacher increased the likelihood that a child would be rated as hyperactive or inattentive regardless of his observed level of activity or attentiveness, indicating that a child’s defiance and disobedience are significant causes of misclassification. This is also known as the halo effect for teachers and it refers to the influence of one class of behaviors into the perception and further ratings of other behaviors (31, 32). Our exploratory factor analysis of teacher ratings revealed also that the items indicating the inattentiveness loaded together in a separate component, suggesting for a separate factor consistent with the arguments of some authors that suggest considering ADHD/I subtype as a distinct diagnostic category (33). Taken together, these differences suggest that certain behaviors might be external manifestations of a range of distinct childhood psychiatric disorders but cannot be considered unique to these disorders.

## 6. LIMITATIONS

Our results must be interpreted with caution in the light of limitations that this study has. First, the sample was not representative of the population of elementary school children and it was selected based on an administrative distribution of elementary schools (schools of only one district of Tirana city were included in the study), indicating the need for further studies in more representative samples of the Albanian youth population. Second, the study was performed in a low risk population and as Goodman (34) suggests when screening for child mental health problems in low-risk epidemiological samples, considering of broader internalizing or externalizing rating scales might be more appropriate. Additionally, feasibility issues need to be further investigated in order to increase the participation and collaboration of school staff for the early screening of mental health problems in youths. Volpe, et al (35), argue that although the efficiency with which a wide range of be-

havioral data can be obtained makes behavior rating scales particularly attractive tools for the purposes of screening and evaluation, feasibility concerns arise in the context of formative assessment. Specifically, informant load, or the amount of time informants are asked to contribute to the assessment process, likely has a negative impact on the quality of data over time and the informant’s willingness to participate.

## 7. CONCLUSION

This is the first study to examine the psychometric properties of the SDQ in a population of Albanian children. Overall these preliminary results support the five dimensions of SDQ with satisfying subscale reliability based on internal consistency coefficients. However, the content of the factors obtained do not fully support the original factor structure proposed by the authors. Because of the suggested non-one-dimensional contribution of several SDQ items, caution is recommended when using the SDQ in Albanian children when screening for Attention Deficit/Hyperactivity Disorder in elementary school population. Further investigations in larger and more representative samples are recommended and establishment of SDQ norms for Albanian youths must be a priority.

- **Author contribution:** The author contributed in all steps of the preparation of this paper.
- **Conflict of interest:** none declared.

## REFERENCES

1. Fornis M, Abad J, Kirchner, T. Internalizing and externalizing problems. In R. J. R. Levesque (Ed.), *Encyclopedia of adolescence*. New York: Springer; 2014: 1464-1469.
2. Lane K. L, Menzies HM, Oakes WP, Kalberg JR. *Systematic screenings of behavior to support instruction: From preschool to high school*. New York: The Guilford Press, 2012.
3. Walker B, Cheney D, Stage S, Blum C, Horner RH. *Schoolwide Screening and Positive Behavior Supports: Identifying and Supporting Students at Risk for School Failure*. *Journal of Positive Behavior Interventions*. 2005; 7(4): 194-204. <https://doi.org/10.1177%2F10983007050070040101>
4. Belfer LM. Child and adolescent mental disorders: The magnitude of the problem across the globe. *Journal of Child Psychology and Psychiatry*. 2008; 49: 226-236. <https://doi.org/10.1111/j.1469-7610.2007.01855.x>
5. Kessler RC, Avenevoli S, Costello EJ, Georgiades K, Green JG, Gruber MJ, et al. Prevalence, persistence, and sociodemographic correlates of DSM-IV disorders in the National Comorbidity Survey Replication Adolescent Supplement. *Archives of General Psychiatry*. 2012; 69: 372-380. <https://doi.org/10.1001/archgenpsychiatry.2011.160>
6. Merikangas KR, He JP, Brody D, Fisher PW, Bourdon K, Koretz DS. Prevalence and treatment of mental disorders among US children in the 2001-2004 NHANES. *Pediatrics*. 2010; 125: 75-81. <https://doi.org/10.1542/peds.2008-2598>
7. Costello EJ, He J, Sampson NA, Kessler RC, Merikangas KR. Services for adolescent psychiatric disorders: 12-month data from the National Comorbidity Survey-Adolescent. *Psychiatric Service*. 2014; 65(3): 359-366. doi: 10.1176/appi.ps.201100518

8. Dowdy E, Ritchey K, Kamphaus RW. School-based screening: a population based approach to inform and monitor children's mental health needs. *School Mental Health*. 2010; 2: 166-176. doi:10.1007/s12310-010-9036-3
9. Severson HH, Walker HM, Hope-Doolittle J, Kratochwill TR, Gresham FM. Proactive, early screening to detect behaviorally at-risk students: Issues, approaches, emerging innovations, and professional practices. *Journal of School Psychology*. 2007; 45(2): 193-223. doi:10.1016/j.jsp.2006.11.003
10. Offord DR, Boyle MH, Racine Y, Szatmari P, Fleming JE, Sanford M, Lipman EL. Integrating assessment data from multiple informants. *Journal of the American Academy of Child and Adolescent Psychiatry*. 1996; 35(8): 1078-1085. doi: 10.1097/00004583-199608000-00019
11. Dvorsky MR, Girio-Herrera E, Owens JS. School-based screening for mental health in early childhood. In M. Weist, N. A. Lever, C. Bradshaw, & J. S. Owens (Eds.), *Handbook of school mental health: Research, training, practice, and policy* (2nd ed.). New York: Springer US. 2014: 297-310 doi:10.1007/978-1-4614-7624-5\_22
12. Girio-Herrera EL, Dvorsky MR, Owens JS. Mental health screening in kindergarten youth: A multistudy examination of the concurrent and diagnostic validity of the Impairment Rating Scale. *Psychological Assessment*. 2015; 27(1): 215-227. <http://psycnet.apa.org/doi/10.1037/a0037787>
13. Goodman, R. The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*. 1997; 38(11): 581-586. doi: 10.1097/00004583-200111000-00015
14. Stone LL, Otten R, Engels RC, Vermulst A, Janssens JM. Psychometric properties of the parent and teacher versions of the Strengths and Difficulties Questionnaire for 4- to 12-Year-olds: A review. *Clinical Child and Family Psychology Review*. 2010; 13: 254-274. doi:10.1007/s10567-010-0071-2
15. Woerner W, Becker A, Rothenberger A. Normative data and scale properties of the German parent SDQ. *European Child & Adolescent Psychiatry*. 2004; 13(2): ii3-ii10. <https://doi.org/10.1007/s00787-004-2002-6>
16. Niclasen J, Teasdale TW, Andersen AMN, Skovgaard AM, Elberling H, Obel C. Psychometric Properties of the Danish Strength and Difficulties Questionnaire: The SDQ Assessed for More than 70,000 Raters in Four Different Cohorts. *PLoS ONE*. 2012; 7(2): e32025. <https://doi.org/10.1371/journal.pone.0032025>
17. Mellor D, Stokes M. The factor structure of the Strengths and Difficulties Questionnaire. *European Journal of Psychological Assessment*. 2007; 23: 105-112. <http://psycnet.apa.org/doi/10.1027/1015-5759.23.2.105>
18. Dickey WC, Blumberg SJ. Revisiting the factor structure of the Strengths and Difficulties Questionnaire: United States, 2001. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2004; 43(9): 1159-1167. doi:10.1097/01.chi.0000132808.36708.a9
19. DeVellis RF. *Scale development: Theory and applications* (2nd Edition). Thousand Oaks, CA: Sage Publications, Inc. 2003; 27-49.
20. Cohen J. *Statistical power analysis for the behavioral sciences* (Rev. ed.). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc., 1977.
21. Briggs SR, Cheek JM. The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality*. 1986; 54: 106-148. <http://dx.doi.org/10.1111/j.1467-6494.1986.tb00391.x>
22. Tabachnick BG, Fidell LS. *Using multivariate statistics*. Needham Heights, MA, Allyn and Bacon, 2001.
23. Byrne BM. *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. London, Lawrence Erlbaum, 2001.
24. Tabachnick BG, Fidell LS. *Using multivariate statistics* (5th Ed.). New York: Pearson, 2007.
25. Horn JL. A rationale and test for the number of factors in factor analysis. *Psychometrika*. 1965; 30: 179-185.
26. Watkins MW. *Monte Carlo PCA for Parallel Analysis* [computer software]. State College, PA, Ed & Psych Associates, 2000.
27. Hill CR, Hughes JN. An examination of the convergent and discriminant validity of the Strengths and Difficulties Questionnaire. *School Psychology Quarterly*. 2007; 22: 380-406.
28. Milich R, Loney J, Landau S. Independent Dimensions of Hyperactivity and Aggression: A Validation With Playroom Observation Data. *Journal of Abnormal Psychology*. 1982; 91(3): 183-198.
29. Hinshaw SP. On the Distinction Between Attentional Deficits/Hyperactivity and Conduct Problems/Aggression in Child Psychopathology. *Psychological Bulletin*. 1987; 101(3): 443-463.
30. Schachar R, Sandberg S, Rutter M. Agreement Between Teachers' Ratings and Observations of Hyperactivity, Inattentiveness, and Defiance. *Journal of Abnormal Child Psychology*. 1986; 14(2): 331-345.
31. Abikoff H, Courtney M, Pelham WE, Koplewicz HS. Teachers' rating of disruptive behaviors: The influence of halo effects. *Journal of Abnormal Child Psychology*. 1993; 21, 519-533.
32. Nisbett RE, Wilson TD. The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*. 1977; 35: 250-256.
33. Roberts W, Milich R. Examining the Changes to ADHD in the DSM-5: One Step Forward and Two Steps Back. *The ADHD Report*. June 2013; 21(4): 1-6. ISSN 1065-8025. © 2013 The Guilford Press, 2013.
34. Goodman A, Lamping DL, Ploubidis GB. 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. *J Abnorm Child Psychol*. 2010; 38(8): 1179-1191.
35. Volpe RJ, Briesch AM, Gadow KD. The efficiency of behavior rating scales to assess inattentive - overactive and oppositional - defiant behaviors: Applying generalizability theory to streamline assessment. *Journal of School Psychology*. 2011; 49: 131-155.