

Validation of the SDQ in a multi-ethnic population of young children

Cathelijne L. Mieloo^{1,2}, MSc, Floor Bevaart³, MSc, Marianne C.H. Donker², Prof., Floor V.A. van Oort³, PhD, Hein Raat², Prof., Wilma Jansen^{1,2}, PhD

¹ The Rotterdam-Rijnmond Public Health Service (GGD Rotterdam-Rijnmond) , Rotterdam, The Netherlands

² Erasmus Medical Centre (Department of Public Health), Rotterdam, The Netherlands

³ Erasmus Medical Centre (Department of Child and Adolescent Psychiatry), Rotterdam, The Netherlands

Word count abstract: 244

Word count text: 3487

Correspondence to:

Wilma Jansen, PhD

Rotterdam-Rijnmond Public Health Service (GGD Rotterdam-Rijnmond)

Department of Youth Policy

PO Box 70032, 3000 LP Rotterdam

Telephone: 31-10-4339968

E-mail: w.jansen@rotterdam.nl

Abstract

Background: The Strengths and Difficulties Questionnaire (SDQ) is a valuable screening tool for identifying psychosocial problems. Its performance in a multi-ethnic society, common to many paediatric healthcare workers, has not been investigated. Because it is important that screening instruments are valid and reliable for all ethnic groups within one society, we examined differences in the SDQ's psychometric properties in a multi-ethnic society.

Methods: The SDQ parent (n=8,114) and teacher form (n=9,355) were completed as part of a preventive health check for children aged 5-6 years of Dutch and non-Dutch ethnic backgrounds. The CBCL/TRF was administered to a subsample.

Results: Factor analysis of the parent-rated SDQ showed different rating patterns for two of the five subscales for non-Dutch children as compared to Dutch children. Cronbach's alpha for the total difficulties score varied by ethnic group (0.73-0.78 parent-rated SDQ, 0.80-0.83 teacher-rated SDQ) and coefficients were generally smaller for non-Dutch than for Dutch children ($p < 0.05$). Alpha coefficients for subscales varied between 0.31-0.85 for ethnic groups. Inter-rater correlations between parents and teachers for the total difficulties score varied between 0.20-0.41 between ethnic groups and were larger for Dutch than for non-Dutch children ($p < 0.05$). Concurrent validity was acceptable for most scales and most ethnic groups.

Conclusion: The total difficulties score of the parent- and teacher-rated SDQ is valid and reliable for different ethnic groups within Dutch society. However, there are differences in reliability and validity of the subscales, which makes interpretation of the subscales difficult for certain ethnic groups.

Keywords: validation, reliability, ethnicity, parent SDQ, teacher SDQ, multi-ethnic society

Introduction

Prevalence of psychosocial problems varies between eight and eighteen percent in young children.¹⁻² Early detection and treatment have an important role in preventing psychosocial problems and may benefit the child's development, well-being, and future health.³ For early detection, professionals in paediatric care need valid and reliable screening instruments. Because societies all over the world are becoming increasingly multi-ethnic and prevalence of psychosocial problems in some minority children is higher than in native children,⁴⁻⁶ it is even more important that these instruments are valid and reliable for all ethnic groups within a society.

The Strengths and Difficulties Questionnaire (SDQ) is a relatively short instrument developed to screen for emotional and behavioural problems in children aged 3-16 years.⁷ It was validated in many countries with satisfying results. The psychometric properties of the SDQ are strong, especially for the teacher version.⁸ However, studies performed in non-western countries showed different reliability and validity outcomes than studies in western countries. Studies of African, Chinese and Arab children indicated only partial agreement with the five-factor structure and certain items did not load on their theoretical factors.⁹⁻¹¹ Furthermore, studies in China and Japan showed lower reliability of the subscales than studies in Great Britain, where the SDQ was developed.¹¹⁻¹² A possible explanation is that parents in non-western countries have different perceptions of deviant behaviour than parents in western countries.¹³ Language and cultural differences in how emotions are expressed could also play a role.¹⁴

Because differences in validity are found between countries, questions arise on the reliability and validity of the SDQ when used in multi-ethnic societies. Two studies reported the factor structure of the SDQ in a multi-ethnic society.¹⁵⁻¹⁶ These studies confirm a similar structure in migrant groups for the self-report and teacher-rated SDQ, but the psychometric properties of the parent-rated SDQ are not yet investigated in groups by ethnic background, although parents are very important informants on the well being of their young child. Furthermore, questions remain about the parent- and teacher-rated SDQ with regard to its internal consistency, inter-rater agreement and construct validity for children of different ethnic groups within one society.

Therefore, in the present study we examined differences in the psychometric properties (factor structure, internal consistency, inter-rater agreement, and concurrent and divergent validity) of the parent and teacher versions of the SDQ by ethnicity of the child. We used data from the regular preventive child healthcare in 5-6 old children living in the Rotterdam-Rijnmond area in the Netherlands. Among these children there are five major ethnic groups with parents who are labour migrants or who come from former Dutch colonies. In the Netherlands, one in four children is of non-Dutch background.¹⁷ This prevalence is even higher in the larger cities, such as in the Rotterdam-Rijnmond area, where one in two children is of non-Dutch background.¹⁷

Methods

Sample and design

The SDQ was administered to parents and teachers as part of the regular preventive child healthcare program for children in grade 2 at elementary school (5 to 6 years of age). A total of 11,987 children, living in the Rotterdam-Rijnmond area in the Netherlands, were eligible for this preventive health check in the school year 2008-2009.

Parents provided questionnaire information on 8,114 (67%) children, and teachers on 9,355 (80%) children. For 6,525 (59.6%) children, both parent and teacher reports were available. Parental non-response was more likely when children had an elevated score on the total difficulties scale of the teacher-rated SDQ (mean (SD) non-responders 6.03 (5.31) versus responders 4.86 (4.79), $\eta^2=0.01$, $p<0.001$). Parental non-response was also more likely when children were of non-Dutch background (non-response among Dutch children 14% versus non-Dutch children 38%, $\eta^2=0.01$, $p<0.001$).

Teacher non-response was more likely when children were of Dutch background (non-response among Dutch children 19% versus non-Dutch children 11%, $\eta^2=0.07$, $p<0.001$), but was independent of the total difficulties score of the parent report ($p=0.81$).

In addition to the SDQ 801 parents filled out the Child Behaviour Checklist (CBCL) and 898 teachers filled out the Teacher Report Form (TRF) for validation purposes.¹⁸ This sample of parents and teachers was selected in two ways. One way was enrolling a random selection of children. This sample received the SDQ and CBCL/TRF at the same time. The other way was enrolling children with an SDQ score above the p90 cut off on the parent and/or teacher-rated SDQ. Parents and teachers received the CBCL/TRF within four weeks of returning the SDQ. For Turkish and Moroccan parents questionnaires with double language (Dutch/Turkish or Dutch/Arabic) were used.

This study was approved by the medical ethics committee of the Erasmus Medical Centre Rotterdam, the Netherlands.

Measures

The SDQ is a 25-item questionnaire with three response categories (not true, somewhat true, and certainly true). The questionnaire has five subscales: emotional symptoms, conduct problems, hyperactivity/inattention problems, peer problems, and prosocial behaviour (Table S1, available online). Each scale consists of 5 items. The summed score of the first four subscales provides a total difficulties score. A high total difficulties score indicates more problems. The prosocial behaviour scale provides information about the child's protective factors. On this scale, a low score indicates more problems. The SDQ was scored in the standard manner, which means that for all children with less than two items missing on a subscale a score was calculated. Further information on the SDQ and scoring is available at www.sdqinfo.com.

The CBCL and the TRF were used to obtain standardized parent and teacher reports of children's problem behaviour. The CBCL and TRF contain 118 problem items with three response categories (not true, somewhat true, and very true or often true). The questionnaire asks about eight empirically based syndromes: Anxious/Depressed, Withdrawn/Depressed, Somatic Complaints, Social Problems, Thought Problems, Attention Problems, Rule-Breaking Behaviour, and Aggressive Behaviour.

Socio-demographic characteristics included gender, age, and country of birth of the child and the child's parents. Irrespective of the child's country of birth, a child's ethnic background was defined as Dutch when both parents were born in the Netherlands. Ethnic background of a child was defined as Surinamese, Antillean/Aruban, Moroccan or Turkish when one or both parents were born in one of these countries. This is according to the definition as used by Statistics Netherlands.¹⁹

Statistical analyses

All analyses were performed with SPSS 19.0 (SPSS Inc. 2010) and repeated separately for each subgroup by ethnic background. Differences in mean scores between parents and teachers were analysed with a paired sample t-test. Differences in mean scores between subgroups by ethnic background were analysed with ANOVA with post hoc test Games Howell because equal variance and equal group sizes were not present.

Principle component analyses with a forced five-factor model were carried out to examine the factor structure of the SDQ. Oblimin rotation was used as correlated factors were hypothesised. A criterion of 0.3 was chosen to reveal cross-loadings.

Internal consistency was determined by means of Cronbach's alpha coefficient. A Cronbach's alpha of at least 0.7 is recommended for instruments intended for use in groups and individuals.²⁰ Differences in Cronbach's alpha by ethnic background were analysed by means of F-statistics.²¹

Inter-rater agreement between parents and teachers was determined with intra-class correlations (ICC) using a two-way random effect model with absolute agreement²² and with Pearson correlations. Differences in correlations by ethnic background were analysed by means of the Fisher R to Z transformation.²³ A Pearson correlation coefficient of 0.27 was considered as normal as this is the meta-analytic mean between parent and teacher reports of emotional and behavioural problems in children.²⁴

Concurrent and divergent validity of the parent- and teacher-rated SDQ was assessed by calculating Pearson's correlation coefficient with the CBCL and the TRF. For concurrent validity, we expected stronger correlations between the SDQ scales and the CBCL scales that rated similar problems. Therefore, the emotional symptoms scale of the SDQ was expected to correlate more strongly with the Internalising scales of the CBCL and TRF than with all other scales. The conduct problems scale and the hyperactivity/inattention scale of the SDQ were expected to strongly correlate with the Externalising scales of the CBCL/TRF, and the peer problems scale of the SDQ was expected to correlate more strongly with the Social Problems scale of the CBCL/TRF as compared to all other scales. For divergent validity we expected negative or no correlations between the SDQ scales and the CBCL/TRF scales rating opposite problems. Therefore, the SDQ prosocial behaviour scale, which contains items about strengths, would have a negative or zero correlation with all scales of the CBCL/TRF, which contain only items about difficulties. The scale emotional symptoms of the SDQ would have a weak correlation with the Externalising subscales of the CBCL and TRF subscales and the SDQ conduct problem scale and the hyperactivity/inattention scale would have a weak correlation with the CBCL/TRF Internalising subscales.

Results

The population consisted of 5,555 boys (51%) and 5,036 girls (49%). Mean age was 5.3 years (Table 1). There were no significant differences in age or gender by ethnic background. There were significant differences in age and ethnicity by rater ($p < 0.001$), but the effect size was small (Cohen's $d = 0.12$, Cramer's $\phi = 0.05$). There were no significant differences in child gender by rater. Parents and teachers reported higher total difficulties scores on almost all subscales in non-Dutch children than in Dutch children. Parents reported more difficulties than teachers on all scales for Dutch and Surinamese children ($p < 0.01$). With the exception of the prosocial behaviour scale, parents of Turkish and Moroccan children reported more difficulties than teachers on all scales (Table S2, available online).

<Table 1>

Factor structure

Principal component analyses of the parent SDQ in Dutch children showed that the first five factors all had Eigen values > 1.0 and accounted for 42.6% of the total variance. For parent ratings, all items loaded on the predicted factors (Table S3, available online). Total variance explained by ethnic group was 42.6% for Surinamese, 46.0% for Antillean/Aruban, 42.1% for Turkish and 41.6% for Moroccan children. Analyses of the parent SDQ for Antillean/Aruban and Turkish children showed that the items on the prosocial behaviour scale, emotional symptoms scale, and hyperactivity/inattention scale loaded on the predicted factors, for Surinamese children only the items of the hyperactivity/inattention scale and prosocial behaviour scale loaded on the predicted factors, and for Moroccan children only the items of the emotional symptoms scale and prosocial behaviour scale loaded on the predicted factors. Items of the peer problems scale and conduct problems scale mainly loaded on the emotional symptoms and prosocial behaviour scale in almost all groups (Table S3, available online).

Principal component analyses of the teacher SDQ for Dutch children showed that the five factors all had Eigen values >1.0 and accounted for 51.5% of the total variance. All items loaded on the predicted factors (Table S4, available online). Total variance explained by ethnic group was 54.7% for Surinamese, 54.6% for Antillean/Aruban, 54.9% for Turkish and 52.4 % for Moroccan children. Analysis of the teacher SDQ for Turkish and Moroccan children showed that several items of the peer problems scale loaded on the prosocial behaviour scale. Analyses of the teacher SDQ for Surinamese and Antillean/Aruban children showed that several items of the conduct problems scale loaded on the hyperactivity/inattention scale and items of the peer problems scale loaded on the emotional problems scale (Table S4, available online).

Internal consistency

Cronbach's alpha's for the total difficulties score and the hyperactivity/inattention scale were above 0.70 for the parent report of Dutch, Surinamese and Antillean/Aruban children (Table 2). For parent reports of Turkish and Moroccan children, only the total difficulties score was above 0.70. Cronbach's alpha's for the hyperactivity/inattention scale on the parent SDQ were higher for Dutch children than for all other ethnic groups. Cronbach's alpha coefficients for the parent report of Moroccan children were generally lower than for Dutch children ($p < 0.05$). Internal consistency did not improve by deleting items.

Cronbach's alpha was above 0.70 for the total difficulties score, emotional symptoms (only Dutch and Turkish children), hyperactivity/inattention and the prosocial behaviour scale in teacher reports (Table 2). Alpha's for the teacher report of Moroccan children for the emotional symptoms scale and hyperactivity/inattention scale were lower than for Dutch children ($p < 0.05$). The difference, however, was small between the ethnic groups (difference α emotional problems 0.08 and hyperactivity/inattention scale 0.02). Internal consistency did not improve by deleting items.

<Table 2>

Inter-rater agreement

ICCs and Pearson correlations were significant for all scales. Pearson correlation coefficients for the total difficulties score and three out of five subscales in Dutch children were larger than the meta-analytic mean of 0.27.²⁴ For Turkish and Moroccan children, only the hyperactivity/inattention scale showed a Pearson correlation coefficient larger than the meta-analytic mean. ICCs of the emotional symptoms scale and the total difficulties scale were significantly larger for Dutch children than all other groups ($p < 0.05$). ICCs of three out of five subscales for Turkish and Moroccan children were smaller than for Dutch children (Table 3).

<Table 3>

Concurrent and divergent validity

The pattern of correlation coefficients for concurrent and divergent validity between the SDQ and CBCL/TRF was as hypothesized for the emotional symptoms, conduct problems, hyperactivity/inattention and prosocial behaviour scale in all groups for the teacher report and in Dutch, Surinamese, Antillean/Aruban, and Turkish children for the parent report (Table 4).

In all groups the peer problems scale showed larger correlations with the Withdrawn/Depressed scale of the CBCL than with the social problems scale of the CBCL. In Surinamese and Moroccan children, the Anxious/Depressed scale of the CBCL was only moderately correlated with the emotional behaviour scale of the SDQ. Further, in Moroccan children the conduct problem scale showed larger correlations with the Withdrawn/Depressed scale than the externalizing CBCL scales.

The emotional symptoms scale of the teacher SDQ has small correlations with the Somatic Complaints subscale of the TRF for all groups except for Moroccan children. Further, in Antillean/Aruban and Turkish children the peer problems scale showed, next to the hypothesized Social Problems scale, substantial correlations with the Anxious/Depressed scale of the TRF.

<Table 4>

Discussion

The present study conducted in a multi-ethnic community sample of young children was the first study, as we know, investigating the psychometric properties (factor structure, inter-rater reliability, internal consistency, and concurrent and divergent validity) of the parent-rated and teacher-rated SDQ for different ethnic groups living in one society. Our findings indicate that although the total difficulties score of the parent and teacher SDQ is valid and reliable for all ethnic groups, there are differences in validity and reliability of the subscales across the different ethnic groups. Further, both versions of the SDQ had higher reliability and validity in Dutch children than in non-Dutch children and the teacher-rated SDQ had higher reliability and validity in all groups than the parent SDQ.

In more detail, our analysis showed that in the non-Dutch groups the five-factor structure was not similar to the hypothesized factors. More specifically, items from the conduct problems scale and the peer problems scale did not load on the hypothesised factors. Closer inspection revealed that the items *lies* and *tempers* on the conduct problems scale of the parent-rated SDQ showed higher loadings on emotional problems in non-Dutch children. This was also seen in studies among African and Chinese children.¹⁰⁻¹¹ It is possible that in non-western countries, certain behaviours are an expression of other emotions than in western countries or that these items are interpreted or valued differently and therefore correlate higher with items from other subscales.^{14, 25} For example, in collective societies children learn to suppress the expression of anger because this is regarded as disrespectful; in individual societies, the expression of anger is seen as assertive behaviour.²⁵ This could also be an explanation for the lower inter-rater agreement in non-Dutch children, because most teachers are of Dutch ethnicity whereas one in two parents are of non-Dutch ethnicity.²⁶ Another explanation could be that the difference in child behaviour at home and at school is more prominent for non-Dutch children. Further, it is also possible that stereotypes and biases can influence the teacher report of emotional and behavioural problems in non-Dutch children,²⁷⁻²⁸ since we found differences in reliability and validity of the teacher-rated SDQ between ethnic groups. This is in agreement with other studies where no or low correlations were found between parental reported psychosocial problems and teacher reported problems among asylum seekers and migrant children.²⁹⁻³⁰ In general the inter-rater agreement

between parent and teacher reports of emotional and behavioural problems in children is low (Pearson $r=0.27$).²⁴ For Turkish and Moroccan children we found somewhat lower agreement. However, all other reliability and the validity measures of the total difficulties score remain satisfactory in all ethnic groups.

Although internal consistency for the total difficulties scale was satisfactory for all groups, this did not account for most subscales of the parent SDQ. These findings are inline with other studies.^{8, 31} Because the scales contain just five items, it should be kept in mind that scales with less items are generally less reliable than scales with more items.³² Although we found some differences in internal consistency by ethnic background, these differences were small.

Finally, the concurrent and divergent validity of the parent- and teacher-rated SDQ were generally acceptable in almost all ethnic groups. However, these analyses included very small groups and should be interpreted with care.

In a previous studies differences factor structure, internal consistency and inter-rater agreement between boys and girls were found.³³⁻³⁵ It is possible that these differences show dissimilarity between the ethnic groups. We therefore have repeated all analyses in subgroups by gender for each group by ethnic background (data not shown). However, dissimilarities were not significant and conclusions about reliability and validity remained the same for boys and girls in subgroups by ethnic background.

It should be acknowledged that the present study has a few shortcomings. First, there was a bias in response. However, the effect size was small. Therefore, we do not expect that non-response influenced the outcomes. Finally, as no measure was included to validate the prosocial behaviour scale, we could not investigate the concurrent validity of this positively phrased subscale.

Our study has several strengths: the large sample of children, reports of multiple informants were available for most children, and our study was conducted in a sample of Dutch, Surinamese, Antillean/Aruban, Moroccan and Turkish children. These ethnic minority groups are also found in large cities in other western European countries. Finally, this sample was derived in the setting of the regular preventive youth health care programme. In other words, questionnaire responses were not anonymous and were used for further care decisions. The outcomes are therefore representative for the

daily practice in the preventive healthcare. However this also means that our findings can not be generalized to an anonymous research setting. Also, our sample is of a specific age group, namely 5 to 6 year old children. Thus, generalizing our findings to an anonymous research setting or children of older age probably needs further research.

The present study generates a number of additional research questions. For example, we found differences in the factor structure of the parent- and teacher-rated SDQ for various ethnic groups. This should be further investigated with confirmatory factor analyses to see if these inconsistencies remain. Further, differences in reliability were found between ethnic groups. To investigate the underlying causes of these differences, item response theory could be applied to investigate if differential item functioning (DIF) is present for specific items.³⁶ Finally, less favourable SDQ scores were found for non-Dutch children. This was also found in other studies, but the question remains if these children really show more problem behaviour or that ratings are just higher for these groups.^{4,6} This could partly be investigated by receiver operating characteristic (ROC) curves and comparison with a clinical sample.

Societies are becoming increasingly multi-ethnic and for the reason that there are differences in validity and reliability between ethnic groups, there are implications for research and for the professionals working in the preventive child health care. As some SDQ subscales have lower reliability compared to cut points seen as acceptable and have an even lower reliability in non-Dutch groups than in Dutch groups, the subscales should be interpreted with care and should only be used as a guideline. Furthermore, the inter-rater reliability is low for non-Dutch groups. For this reason, it is important that professionals consult both parents and teachers when evaluating behaviour of a child from a migrant family.

In conclusion, this study provides further support for the validity and reliability of the total difficulties score of the parent-rated and teacher-rated SDQ for detecting psychosocial problems in children in a multi-ethnic society. The total difficulties score of the parent and teacher SDQ is valid and reliable for different ethnic groups within the Dutch society. However, there are differences in reliability and validity of the subscales between ethnic groups. Especially the lower interrater reliability for certain ethnic groups can make interpretation of the SDQ subscales more difficult.

Therefore we only recommend the use of the total difficulties score for screening purposes. Further investigation is needed to understand the underlying causes for these differences.

Acknowledgements

This study was supported by the Netherlands Organisation for Health Research and Development (ZON-MW), grant number 157001017.

Conflicts of interest

None declared.

Keypoints

- The SDQ is suitable as a screening instrument for use in the preventive child health care in a multi-ethnic society.
- The total difficulties score of the parent and teacher SDQ is valid and reliable for all ethnic groups in a multi-ethnic society.
- The subscales of the SDQ should be interpreted with care especially in children with a non-western ethnic background.
- We recommend using multiple informants in order to reduce bias by informants.

References

1. Kelleher KJ, McInerney TK, Gardner WP, Childs GE and Wasserman RC. Increasing identification of psychosocial problems: 1979-1996. *Pediatrics* 2000;105:1313-21
2. Sourander A. Emotional and behavioural problems in a sample of Finnish three-year-olds. *Eur Child Adolesc Psychiatry* 2001;98-104
3. Licence K. Promoting and protecting the health of children and young people. *Child Care Health and Dev* 2004;30:623-635
4. Bengi-Arslan L, Verhulst FC, Ende J and Erol N. Understanding childhood (problem) behaviors from a cultural perspective: comparison of problem behaviors and competencies in Turkish immigrant, Turkish and Dutch children. *Soc Psychiatry Psychiatr Epidemiol* 1997;V32:477-484
5. Stevens GW and Vollebergh WAM. Mental Health in migrant children. *J Child Psychol Psychiatry* 2008;49:276-294
6. Horwitz SM, Leaf PJ and Leventhal JM. Identification of psychosocial problems in pediatric primary care: do family attitudes make a difference? *Arch Pediatr Adolesc Med* 1998;152:367-71
7. Goodman R. The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol Psychiatry* 1997;38:581-6
8. Stone LL, Otten R, Engels RC, Vermulst AA and Janssens JM. Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-year-olds: a review. *Clin Child Fam Psychol Rev* 2010;13:254-74
9. Thabet AA, Stretch D and Vostanis P. Child mental health problems in Arab children: application of the strengths and difficulties questionnaire. *Int J Soc Psychiatry* 2000;46:266-80
10. Kashala E, Elgen I, Sommerfelt K and Tylleskar T. Teacher ratings of mental health among school children in Kinshasa, Democratic Republic of Congo. *Eur Child Adolesc Psychiatry* 2005;14:208-15
11. Du Y, Kou J and Coghill D. The validity, reliability and normative scores of the parent, teacher and self report versions of the Strengths and Difficulties Questionnaire in China. *Child Adolesc Psychiatry Ment Health* 2008;2:8

12. Matsuishi T, Nagano M, Araki Y, et al. Scale properties of the Japanese version of the Strengths and Difficulties Questionnaire (SDQ): a study of infant and school children in community samples. *Brain Dev* 2008;30:410-5
13. Weisz JR, Suwanlert S, Chaiyasit W, Weiss B, Walter BR and Anderson WW. Thai and American perspectives on over- and undercontrolled child behavior problems: exploring the threshold model among parents, teachers, and psychologists. *J Consult Clin Psychol* 1988;56:601-9
14. Russell JA. Culture and the categorization of emotions. *Psychol Bull* 1991;110:426-50
15. Richter J, Sagatun A, Heyerdahl S, Oppedal B and Roysamb E. The Strengths and Difficulties Questionnaire (SDQ) - Self-Report. An analysis of its structure in a multiethnic urban adolescent sample. *J Child Psychol Psychiatry* 2011;
16. Zwirs BW, Burger H, Schulpen TW, Vermulst AA, Hira Sing RA and Buitelaar JK. Teacher Ratings of Children's Behavior Problems and Functional Impairment Across Gender and Ethnicity: Construct Equivalence of the Strengths and Difficulties Questionnaire. *Journal of Cross-Cultural Psychology* 2011;42:466-481
17. CBS, www.cbs.nl Accessed [2008-2009]
18. Achenbach TM and Rescorla LA. Manual for the ASEBA School-Age Forms & Profiles. Burlington (VT): University of Vermont, Research Center for Children, Youth & Families, 2001
19. CBS, <http://www.cbs.nl/en-GB/menu/themas/dossiers/allochtonen/methoden/begrippen/default.htm?Languageswitch=on&ConceptID=37> Accessed
20. Nunnally JC and Bernstein IH. *Psychometric theory*. 3rd. New York: McGraw-Hill, 1994
21. Glantz SA. *Primer of Biostatistics*. 3rd. New York: McGraw-Hill, 1992
22. Field A. Intraclass correlation. In: D. H. B. Everitt. *Encyclopedia of Behavioral Statistics*. New York: Wiley, 2005: 948-954
23. Corey DM, Dunlap WP and Burke MJ. Averaging Correlations: Expected Values and Bias in Combined Pearson r s and Fisher's z Transformations. *J Gen Psychol* 1998;125:245-261

24. Achenbach TM, McConaughy SH and Howell CT. Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. *Psychol Bull* 1987;101:213-32
25. Stephan WG, Stephan CW and Cabezas de Vargas M. Emotional expression in Costa Rica and the United States. *J Cross Cult Psychol* 1996;27:147-160
26. CBS,
<http://statline.cbs.nl/StatWeb/publication/?DM=SLEN&PA=71822ENG&D1=1&D2=0&D3=0&D4=1-2&D5=0&D6=2&D7=7-8&LA=EN&VW=T> Accessed [2011]
27. Tenenbaum HR and Ruck MD. Are teachers' Expectations Different for Racial Minority Than for European American Students? A Meta-Analysis. *J Educ Psychol* 2007;99:253-273
28. Dusek JB and Joseph G. The bases of teacher expectancies: A meta-analysis. *J Educ Psychol* 1983;327-346
29. Vollebergh WA, ten Have M, Dekovic M, et al. Mental health in immigrant children in the Netherlands. *Soc Psychiatry Psychiatr Epidemiol* 2005;40:489-96
30. Wiegiersma PA, Stellinga-Boelen AA and Reijneveld SA. Psychosocial problems in asylum seekers' children: the parent, child, and teacher perspective using the Strength and Difficulties Questionnaire. *J Nerv Ment Dis* 2011;199:85-90
31. Hölling H, Erhart M, Ravens-Sieberer U and Schlack R. Behavioural problems in children and adolescents. First results from the German Health Interview and Examination Survey for Children and Adolescents (KiGGS). *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2007;50:784-93
32. Streiner DL and Norman GR. *Health Measurement Scales*. Oxford: Oxford University Press, 1989
33. Hawes DJ and Dadds MR. Australian data and psychometric properties of the Strengths and Difficulties Questionnaire. *Aust N Z J Psychiatry* 2004;38:644-51
34. Hill CR and Hughes JN. An Examination of the Convergent and Discriminant Validity of the Strengths and Difficulties Questionnaire. *Sch Psychol Q* 2007;22:380-406

35. Mieloo C, Raat H, van Oort F, et al. Validity and reliability of the strengths and difficulties questionnaire in 5-6 year olds: differences by gender or by parental education? PLoS One 2012;7:e36805

36. Reise SP, Ainsworth AT and Haviland MG. Item Response Theory: Fundamentals, Applications, and Promise in Psychological Research. Curr Dir Psychol Sci 2005;14:95-101

Table 1 **Population characteristics**

Questionnaire	Parent-rated forms		Teacher-rated forms	
	SDQ (n=8114)	CBCL (n=801)	SDQ (n=9355)	TRF (n=898)
Mean age of child (SD)	5.3 (0.51)	5.2 (0.5)**	5.3 (0.51)	5.2 (0.4)**
Gender of child (male)	4107 (50.7%)	402 (55.5%)**	4778 (51.3%)	449 (51.7%)
Ethnicity of child*				
Dutch	4750 (58.6%)	350 (49.9%)**	4516 (53.0%)	498 (62.2%)**
Surinamese	521 (6.4%)	58 (8.3%)**	619 (7.3%)	60 (7.5%)**
Antillean/Aruban	264 (3.8%)	34 (4.8%)**	339 (4.0%)	21 (2.6%)**
Turkish	661 (8.2 %)	69 (9.8%)**	759 (8.9%)	36 (4.5%)**
Moroccan	623 (7.7%)	46 (6.6%)**	811 (9.5%)	55 (6.9%)**
Other	1281 (15.8%)	145 (20.6%)**	1481 (17.4%)	131 (16.3%)**

*= significant difference between parent and teacher report (p<0.05)

**= significant difference between SDQ and ASEBA form (p<0.05)

Table 2 Internal consistency by ethnic background**Parent report (Cronbach alpha)**

	Dutch	Surinamese	Antillean/Aruban	Turkish	Moroccan
	n=4384	n=460	n=200	n=537	n=480
Emotional symptoms	0.61	0.52*	0.50*	0.60	0.58
Conduct problems	0.51	0.51	0.55	0.48	0.44*
Hyperactivity/inattention	0.78	0.74*	0.72*	0.67*	0.65*
Peer problems	0.49	0.50	0.51	0.31*	0.35*
Prosocial behaviour	0.63	0.58	0.68	0.61	0.60
Total difficulties score	0.77	0.77	0.78	0.77	0.73*

Teacher report (Cronbach alpha)

	n=4342	n=596	n=322	n=739	n=783
Emotional symptoms	0.71	0.65	0.69	0.73	0.67*
Conduct problems	0.60	0.68	0.66	0.64	0.66
Hyperactivity/inattention	0.85	0.84	0.85	0.84	0.83*
Peer problems	0.56	0.62	0.58	0.59	0.53
Prosocial behaviour	0.81	0.82	0.82	0.83	0.82
Total difficulties score	0.81	0.82	0.83	0.82	0.80

* = significant difference with Dutch sample ($p < 0.05$)

Table 3 Inter-rater agreement for SDQ scores Parent x Teacher

ICC (Pearson)

	Ethnicity of child				
	Dutch	Surinamese	Antillean/ Aruban	Turkish	Moroccan
SDQ	n=3718	n=435	n=207	n=535	n=516
Emotional Symptoms	0.28 (0.29)	0.11 (0.11)*	0.11 (0.12)*	0.13 (0.14)*	0.09 (0.10)*
Conduct problems	0.23 (0.25)	0.22 (0.23)	0.27 (0.28)	0.17 (0.19)	0.16 (0.17)
Hyperactivity/ inattention	0.42 (0.45)	0.41 (0.42)	0.40 (0.41)	0.31 (0.32)*	0.29 (0.29)*
Peer problems	0.29 (0.29)	0.23 (0.24)	0.22 (0.23)	0.18 (0.21)*	0.08 (0.09)*
Prosocial behaviour	0.21 (0.22)	0.12 (0.14)	0.32 (0.32)*	0.18 (0.18)	0.12 (0.13)
Total difficulties score	0.41(0.41)	0.28 (0.30)*	0.32 (0.35)*	0.23 (0.26)*	0.20 (0.22)*

* = significant difference with Dutch sample (p<0.05)

Table 4 Concurrent and divergent validity by ethnic background

		Ethnicity of child				
		Dutch	Surinamese	Antillean/ Aruban	Turkish	Moroccan
SDQ scale	CBCL/TRF scale					
Parent report		n=344	n=54	n=29	n=63	n=38
Emotional symptoms	Internalizing problems	0.62	0.49	0.68	0.58	0.49
	Externalising problems	0.36	0.27	0.56	0.43	0.36
	Anxious/ Depressed	0.60	0.38*	0.65	0.63	0.27*
Conduct problems	Internalizing	0.25	0.32	0.51	0.26	0.27
	Externalising	0.60	0.55	0.47	0.45	0.24*
	Withdrawn/ Depressed	0.27	0.39	0.33	0.23	0.33
Hyperactivity/ inattention	Internalising	0.26	0.41	0.33	0.13	0.38
	Externalising	0.47	0.65	0.29	0.34	0.62
Peer problems	Social problems	0.47	0.37	0.53	0.10*	0.04
	Withdrawn/ Depressed	0.48	0.42	0.57	0.14*	0.18

Prosocial	Internalizing	-0.09	-0.03	-0.50*	-0.15	-0.12
	Externalising	-0.28	-0.24	-0.53	-0.02	-0.11
Teacher report		n=516	n=60	n=21	n=37	n=58
Emotional symptoms	Internalizing	0.64	0.47	0.34	0.73	0.72
	Externalising	0.09	0.23	0.16	-0.30*	0.10
	Somatic Complaints	0.22	0.23	0.30	0.35	0.53*
Conduct problems	Internalizing	0.05	-0.04	0.29	-0.23	0.20
	Externalising	0.68	0.63	0.57	0.82	0.82*
Hyperactivity/inattention	Internalizing	0.16	0.15	0.31	0.04	-0.05
	Externalising	0.45	0.56	0.60	0.38	0.16*
Peer problems	Social	0.43	0.32	0.59	0.53	0.49
	Anxious/Depressed	0.23	0.09	0.61*	0.58*	0.40
Prosocial	Internalizing	-0.23	-0.20	-0.71*	-0.18	-0.13
	Externalising	-0.44	-0.24	-0.52	-0.52	-0.46

* = significant difference with Dutch sample (p<0.05)