University of Groningen

Teacher Ratings of Children's Behavior Problems and Functional Impairment Across Gender and Ethnicity

Zwirs, Barbara; Burger, Huibert; Schulpen, Tom; Vermulst, A. A.; HiraSing, R. A.; Buitelaar, Jan

Published in:
Journal of Cross-Cultural Psychology

DOI:
[10.1177/0022022110362752](https://doi.org/10.1177/0022022110362752)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2011

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Zwirs, B., Burger, H., Schulpen, T., Vermulst, A. A., HiraSing, R. A., & Buitelaar, J. (2011). 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*, 42(3), 466-481. <https://doi.org/10.1177/0022022110362752>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Teacher Ratings of Children's Behavior Problems and Functional Impairment Across Gender and Ethnicity: Construct Equivalence of the Strengths and Difficulties Questionnaire

Barbara Zwirs, Huibert Burger, Tom Schulpen, A.A. Vermulst, R.A. HiraSing and Jan Buitelaar

Journal of Cross-Cultural Psychology 2011 42: 466 originally published online 16 July 2010

DOI: 10.1177/0022022110362752

The online version of this article can be found at:

<http://jcc.sagepub.com/content/42/3/466>

Published by:

<http://www.sagepublications.com>

On behalf of:

[International Association for Cross-Cultural Psychology](#)

Additional services and information for *Journal of Cross-Cultural Psychology* can be found at:

Email Alerts: <http://jcc.sagepub.com/cgi/alerts>

Subscriptions: <http://jcc.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://jcc.sagepub.com/content/42/3/466.refs.html>

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
42(3) 466–481
© The Author(s) 2011
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0022022110362752
jccp.sagepub.com

Barbara Zwirs^{1,2}, Huibert Burger^{2,3}, Tom Schulpen⁴,
A. A. Vermulst⁵, R. A. HiraSing⁶, and Jan Buitelaar⁷

Abstract

The present study examined construct equivalence of the teacher Strengths and Difficulties Questionnaire and compared mean scores in an ethnically diverse sample of children living in the Netherlands. Elementary schoolteachers completed the Strengths and Difficulties Questionnaire for 2,185 children aged 6 to 10 years of the four largest ethnic groups in the Netherlands, namely native Dutch ($n = 684$) and Moroccan ($n = 702$), Turkish ($n = 434$), and Surinamese ($n = 365$) immigrant children. Multigroup confirmatory factor analysis suggested the factor structure of the Strengths and Difficulties Questionnaire to be invariant across children's ethnicity and gender. Additionally, the factor structure appeared to be similar for Dutch and Surinamese teachers. However, mean scores on emotional problems, hyperactivity, conduct problems, prosocial behavior, and impairment varied significantly according to ethnicity and gender. Mean scores on peer problems differed significantly for boys and girls, but not across ethnicity. Whether mean differences reflect a method bias or actual differences in classroom behaviors is discussed and needs further research.

Keywords

Strengths and Difficulties Questionnaire, SDQ, construct equivalence, impairment, ethnicity, gender

¹Erasmus University Rotterdam, Erasmus School of Law, Rotterdam, the Netherlands

²University Medical Centre Utrecht, Utrecht, the Netherlands

³University Medical Centre Groningen, University of Groningen, Groningen, the Netherlands

⁴Professor Emeritus, Utrecht University, Utrecht, the Netherlands

⁵Behavioural Science Institute of the University of Nijmegen, Nijmegen, the Netherlands

⁶VU University Medical Centre, Amsterdam, the Netherlands

⁷Radboud University Nijmegen Medical Center, Nijmegen, the Netherlands

Corresponding Author:

Barbara Zwirs, Erasmus University Rotterdam, Erasmus School of Law, Department of Criminology,
Room L7-21, PO Box 1738, 3000 DR Rotterdam, the Netherlands
Email: Zwirs@frg.eur.nl

Comparing the prevalence of childhood psychopathology is of great importance because it may increase our understanding of the cause of psychopathology and because epidemiologic data offer a scientific basis for development of preventive interventions and treatment programs. Although the majority of cross-cultural studies have compared children across different countries (Bird, 1996; Crijnen, Achenbach, & Verhulst, 1999; Roberts, Attkisson, & Rosenblatt, 1998; Verhulst et al., 2003), cross-cultural studies that examine cross-cultural differences in children who live in the same country are increasing in number (Epstein, March, Conners, & Jackson, 1998; Kvernmo & Heyerdahl, 1998; Reid, Casat, Norton, Anastopoulos, & Temple, 2001; Stevens et al., 2003; Zwirs, Burger, Schulpen, & Buitelaar, 2006). Given findings of studies in the United States (Kataoka, Zhang, & Wells, 2002; Olfson, Gameroff, Marcus, & Jensen, 2003; Rowland et al., 2002; Safer & Malever, 2000) and in the Netherlands (Zwirs, Burger, Schulpen, & Buitelaar, 2006) that treatment rates are lower for immigrant children than for native-born children, cross-cultural studies among ethnically diverse children who reside in the same country is of particular relevance.

In the Netherlands, mean differences in problem behaviors have been reported for children from the four largest ethnic groups residing in the Netherlands, namely native Dutch youth and Moroccan, Turkish, and Surinamese immigrant youth (Bengi-Arslan, Verhulst, Van der Ende, & Erol, 1997; Murad, Joung, van Lenthe, Bengi-Arslan, & Crijnen, 2003; Stevens et al., 2003; Zwirs, Burger, Schulpen, & Buitelaar, 2006). However, it remains unclear which explanations account for the observed differences. First, the reported differences may be explained by rater bias. In a pioneering study, Sonuga-Barke & Minocha (1993) showed that English teachers rated Asian children higher on hyperactivity symptoms than English native children, whereas actometer readings and behavioral observations were similar for both groups. In a study by Crijnen, Bengi-Arslan, and Verhulst (2000), native Dutch children and Turkish children were rated by Dutch teachers as well as Turkish immigrant teachers, enabling a comparison of Turkish and Dutch teachers' ratings. It appeared that Turkish immigrant children obtained higher scores on problem behaviors from Turkish immigrant teachers than from Dutch teachers.

Second, the described differences in behavior may reflect actual differences in behavior. For instance, Weisz, Chaiyasit, Weiss, Eastman, and Jackson (1995) reported differences in observations of problem behavior between Thai and American schoolchildren. Likewise, ethnic differences in behavioral observations have been demonstrated for Caucasian and African American schoolchildren (Epstein et al., 2005).

Third, the reported differences may also be due to biased measurement. If an instrument does not measure similar theoretical constructs across groups (i.e., there is no construct equivalence), then differences in mean scores may be caused by differences in underlying constructs. Therefore, prior to the examination of behavior problems in different ethnic groups, the cross-cultural equivalence of the factor structure should be established (Berry, Poortinga, Segall, & Dasen, 2002).

Cross-cultural construct equivalence can exist at different levels (Steenkamp & Baumgartner, 1998). *Configural invariance* is supported when the same symptoms are used to classify a construct, implying that the pattern of zero and nonzero loadings is similar across ethnicity. A more stringent test of equivalence is that of *metric invariance*, called "weak metric invariance" (Meredith, 1993), which implies that the factor loadings on the construct are the same for the different ethnic groups. Whereas configural invariance and metric invariance only require information about the covariation of the items in the different ethnic groups, *scalar invariance* or "strong metric invariance" (Meredith, 1993) additionally evaluates whether the indicator intercepts are the same across the different ethnic groups. Hence, scalar invariance implies that cross-ethnic differences in the means of the observed items are caused by differences in the means of the underlying concepts. Strictly, only after this level of invariance is established can one make meaningful

comparisons of factor means (i.e., normative equivalence) (Steenkamp & Baumgartner 1998). In addition, Van de Vijver and Leung (2001) distinguish *functional equivalence*, which means that the measured constructs show “similar internal structures and similar relationships with other variables” (Van de Vijver & Leung, 2001, p. 1015).

Previous studies that tested the equivalence of behavior problems across ethnicity were mainly focused on configural invariance, using single samples of children. These studies yielded evidence for the cross-cultural equivalence of behavior problems (Dumenci, Erol, Achenbach, & Simsek, 2004) as well as for the nonequivalence across cultures (Kashala, Elgen, Sommerfelt, & Tylleskar, 2005; Thabet, Stretch, & Vostanis, 2000). For instance, the factor structure of the Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1994, 1997) was not confirmed in a sample of Arab children living in the Gaza strip (Thabet et al., 2000) because certain items (i.e., distractibility, feeling scared, feeling unhappy, stealing) did not load highly on the proposed factors. The factor structure of the SDQ was neither supported in children from Kinshasa, Democratic Republic of Congo (Kashala et al., 2005), because most of the items did not have their highest loadings on the expected factors. These two studies suggest the underlying constructs to have a different meaning in an Arab and an African population. In contrast, the five-factor structure of the SDQ did emerge in a Norwegian sample (Ronning, Handegaard, Sourander, & Mørch, 2004) and a German sample (Woerner, Becker, & Rothenberger, 2004). Likewise, the eight-factor structure of the Child Behavior Checklist (CBCL) has been supported in a Turkish population (Dumenci et al., 2004) and in several other countries (Ivanova et al., 2007) but was questioned in a study using different estimation methods (Hartman et al., 1999).

Studies that examined scalar and/or metric equivalence of behavior disorders in children of different ethnic origin simultaneously are less numerous and have predominantly been carried out in the United States (Beiser, Dion, & Gotowiec, 2000; Epstein et al., 1998; Reid et al., 2001; Reid et al., 1998). These studies lend some support for the cross-cultural equivalence of the externalizing problems as measured by the Conners Teacher Rating Scale and of externalizing as well as internalizing problems as assessed by the CBCL. For instance, the subscales of the Conners Teacher Rating Scale have been found to be similar for African American and White American children, with the exception that an additional Antisocial factor was observed in Black boys and an Inattention factor in White girls (Epstein et al., 1998). The Inattention/Overactivity with Aggression (IOWA) Conners Scale has been found to be similar for African American and Caucasian American children (Reid et al., 2001). Likewise, Gross et al. (2006) reported the Externalizing and Internalizing Scales of the CBCL to be generally similar for African American, Latino, and non-Latino White parents, with the exception that some internalizing items functioned differently according to ethnicity (Gross et al., 2006).

The present study examines the construct equivalence and mean scores of internalizing and externalizing problems, as assessed by teachers using the SDQ, in native Dutch children and Moroccan, Turkish, and Surinamese immigrant children living in the Netherlands. More specifically, we addressed (a) whether construct equivalence of the SDQ is supported across the child's ethnicity and gender, (b) whether construct equivalence of the SDQ is supported across the teacher's ethnicity, and (c) whether the relationship between the symptoms scales and the impairment scale is similar across ethnicity (i.e., functional equivalence), given that *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)* criteria include symptoms as well as functional impairment.

The current study differs from former studies in several ways. First, we tested not only for configural invariance but also for metric and scalar invariance. Second, to our knowledge, this study is the first to assess the construct equivalence model of internalizing and externalizing problems in a sample including Moroccan and Surinamese (immigrant) children. Third, we simultaneously tested the equivalence in a sample of four ethnic groups residing in the same

country. Fourth, we assessed not only the construct equivalence but also the relationship between symptom scales and functional impairment across ethnicity (i.e., functional equivalence).

Methods

Participants

Data are taken from the Detection of ADHD among children of Different Ethnic Origins in the Netherlands Study (ADEON), which was carried out in 2002 and 2003 in two large cities (Amsterdam and Utrecht) in the Netherlands. The aim of ADEON was to examine the prevalence of externalizing disorders and service use in children of the four largest ethnic groups in the Netherlands.

To obtain a sample with all four ethnicities represented, we sampled schools from low socioeconomic status (SES) areas (Knol, 1998) with a large immigrant population. Of 87 eligible schools, 45 schools (52%) agreed to participate. In 37 of the 45 participating schools (82%) all teachers participated, the majority being of Dutch origin (77%). Nonparticipating schools and teachers refused to take part because of logistic reasons. However, we assume the sociodemographic characteristics of the children in the 45 participating schools to be similar to those in the 42 nonparticipating schools, because both participating and nonparticipating schools are from the same neighborhoods, which are characterized by a low SES level and a large minority population.

Parents of all 2,802 children enrolled in grades 3 through 5 of the participating schools were asked permission for their child's teacher to administer a screening questionnaire. Children with an ethnic origin other than Dutch, Moroccan, Turkish, or Surinamese were excluded from the study ($n = 336$). As 281 (11%) parents refused to give permission, teachers completed 2,185 of 2,466 questionnaires (89%), with no significant variation in this proportion across ethnicity.

Following the standard classification of the Dutch Central Bureau of Statistics, children were categorized as Moroccan, Turkish, or Surinamese when the child himself or herself or at least one parent had been born in Morocco, Turkey, or Surinam, respectively. When both parents were of non-Dutch origin, we used the mother's country of birth to determine the child's ethnicity. Native Dutch children were Dutch-born offspring of Dutch-born parents. Of the 2,185 subjects included in the study, 31% were of Dutch, 32% of Moroccan, 20% of Turkish, and 17% of Surinamese origin.

The study protocol was approved by the Medical Ethical Committee of University Medical Center Utrecht. Parents of all participating children gave informed consent.

Instrumentation

Teachers completed the Dutch version of the SDQ (Goodman, 1997; Van Widenfelt, Goedhart, Treffers, & Goodman, 2003). The SDQ has been translated into more than 40 languages (see www.sdqinfo.com) and explored in several cultures (Kashala et al., 2005; Thabet et al., 2000; Woerner, Fleitlich-Bilyk, Martinussen, et al., 2004). The reliability of the SDQ has been found acceptable (Goodman, 2001), and the predictive validity has been established in a clinical sample (Goodman, Renfrew, & Mullick, 2000) as well as in a community sample (Goodman, Ford, Simmons, Gatward, & Meltzer, 2000). Recently, the predictive validity of some SDQ items was ascertained in the current sample (Zwirs, Burger, Schulpen, & Buitelaar, 2008). The SDQ includes 25 items describing positive and negative attributes of children. These 25 items are distributed among five scales of five items each: Emotional Symptoms, Conduct Problems, Hyperactivity/Inattention, Peer Problems, and Prosocial Behavior. Each item is scored on scale whereby 0 = *not true*, 1 = *somewhat true*, and 2 = *certainly true*.

The SDQ includes four items that ask about impairment (i.e., distress, interference with friendships, interference with learning, and chronicity) rated on a 4-point scale from 0 (*not at all*) to 3 (*very much*). After we excluded the item on chronicity, because most teachers knew their pupils for only 6 months, the remaining three impact impairment items were summed to create an impairment score (minimum = 0, maximum = 9). Because 80% of the children scored zero on the impairment score, we decided to dichotomize the score at this value. Consequently, functional impairment was defined by a score greater than zero.

The SDQ has been translated and validated for Dutch children, which yielded a factor solution for the parent and youth SDQ that was nearly identical with the five factors as proposed by Goodman (Muris, Meesters, & Van Den, 2003). Widenfelt et al. (2003) reported Cronbach's alpha coefficients greater than .70 for all subscales of the Dutch teacher SDQ and higher interinformant correlations than for the corresponding Achenbach scales (Achenbach & Rescorla, 2001).

Information about SES was obtained by an area-based method. Because the postal code has been related to income, level of education, and rate of unemployment (Knol, 1998), it may be considered as a proxy measure for SES. In the present study, SES was assessed by using the postal code of schools and scored on a 5-point scale from 1 (*high*) to 5 (*low*). These codes were obtained from Statistics Netherlands.

Analysis

The construct equivalence of the SDQ was examined by testing the equality of the factor structure across child ethnicity, child gender, and teacher ethnicity. For teacher ethnicity it was only possible to make a comparison between Dutch ($n = 1,830$) and Surinamese teachers ($n = 254$), because Moroccan and Turkish teachers were a minority, with only 17 (0.8%) and 26 (1.2%) teachers, respectively. We used the software package MPLUS, version 4.2 (Muthén & Muthén, 1998-2006) and Statistical Package for the Social Sciences, version 15.0.

Because the measurement level of the scores on the SDQ items is more ordered-categorical (ordinal) than interval, we used the weighted least squares estimator with a mean and variance adjusted chi-square statistic (WLSMV). This estimator is especially developed for ordered-categorical variables (Muthén & Muthén, 1998-2006). In this case, the ordered-categorical variables are replaced by underlying latent variables having a probability curve derived from the normal distribution. The three categories of each variable are replaced by two thresholds. The adjusted (or scaled or robust) chi-square statistic is derived by dividing the normal-theory chi-square by a scaling correction to better approximate chi-square under nonnormality (<http://www.statmodel.com/chidiff.shtml>). This can explain why a robust chi-square (and other fit measures based on chi-square) can be lower in constrained models in comparison to unconstrained models (see the Results section). Because children are nested within classes within schools, the data have a multilevel structure. Therefore, the data cannot be considered independent. To correct for the nonindependence (complexity) of the data, the COMPLEX procedure in MPLUS was used. With this procedure, the standard errors of the parameter estimates are corrected to obtain unbiased estimates at the class level.

Model fit is given in robust chi-square variates with estimated degrees of freedom. Because the number of degrees of freedom is dependent on sample information, it can vary across identical models (Muthén & Muthén, 1998-2006). Together with the robust chi-square value, model fit is given with two descriptive fit measures: the Comparative Fit Index (CFI; Bentler, 1990) and the root mean square error of approximation (RMSEA; Byrne, 1998). Values of CFI above .95 are preferred (Kaplan, 2000) but should not be lower than .90 (Kline, 1998). Values of RMSEA lower than .05 are preferable but values between .05 and .08 are indicative of fair fit (Joreskog & Sorbom, 1993). However, according to Marsh, Hau, and Wen (2004, p. 325),

conventional goodness-of-fit criteria in confirmatory factor analysis can be too restrictive. Marsh et al. report that it is almost impossible to get an acceptable fit (e.g., CFI > .90, RMSEA < .05). Chen, Curran, Bollen, Kirby, and Paxton (2008) conclude "that there is no empirical support for the use of .05 or any other value as universal cutoff values to determine adequate model fit" (2008, p. 476).

Following Steenkamp and Baumgartner (1998), we first tested for configural invariance, which implies that the pattern of zero loadings and nonzero (or near-zero) loadings of the items shows the same configuration in all four ethnic groups. The configural invariance model, in which factor loadings and thresholds are free to vary, is the baseline model against which the following constrained model will be compared. Second, we assessed a more stringent test of invariance, the metric invariance and the scalar invariance (Steenkamp et al., 1998). As already noted, the measurement level of the factor indicators is ordered-categorical. With this kind of variables, testing of metric and scalar invariance must be done by constraining factor loadings and thresholds simultaneously, because the item probabilities of the factor indicators are influenced by both types of parameters (Muthén & Muthén, 1998-2006).

The metric and scalar equivalence of the five-factor model was tested by comparing the fit of the baseline model with the model where factor loadings and thresholds are constrained to be equal across groups. A chi-square difference test was used to assess whether the constraints significantly deteriorated the model. In MPLUS, the robust chi-square values and the number of degrees of freedom of the baseline model and the constrained model are rescaled to standard chi-square values and degrees of freedom to compute a correct chi-square difference with a correct number of degrees of freedom. Given the large sample size, even small differences could become statistically significant, however. As is noted in literature, "The difference in χ^2 -difference test applied to nested models has essentially the same strengths and weaknesses as the χ^2 -test applied to any single model, namely, the test is directly affected by sample size, and for large samples trivial differences may become significant" (Schermele-Engel, Moosbrugger, & Müller, 2003, p. 34). Therefore, we examined statistical significance, but we also assessed the fit indices (CFI and RMSEA) from the constrained and the unconstrained model.

We conducted two-way analyses of variance and Bonferroni-adjusted post hoc comparisons, controlling for SES and dependency of the observations within classes, to assess whether mean scores varied by child's ethnicity and gender.

Finally, examining functional equivalence (van de Vijver & Leung, 2001), we used logistic regression to estimate the relative risk of the five factors for impairment in each ethnic group, separately adjusting for SES. Relative risks were expressed as odds ratios with 95% confidence intervals.

Results

Demographic characteristics of the study population are presented in Table 1. Irrespective of ethnic group, the proportion boys was about 50%, the mean age was almost 8 years, and SES was low. The percentage of children born in the country of origin was small across all immigrant groups but slightly larger among Surinamese than among Moroccans and Turks. Most immigrant parents (91.7% of mothers and 89.3% of fathers) were born in the country of origin, and the percentage of interethnic relationship was low across all immigrant groups (9.8%).

Construct Equivalence

Configural invariance was assessed by testing the original five-factor model allowing factor loadings and thresholds free to vary (Model 1). As shown in Table 2, the five-factor solution

Table 1. Demographic Characteristics of the Study Sample

	Dutch (<i>n</i> = 684)	Moroccan (<i>n</i> = 702)	Turkish (<i>n</i> = 434)	Surinamese (<i>n</i> = 365)
Boys, <i>n</i> (%)	352 (51.5)	378 (53.8)	212 (48.8)	168 (46)
Girls, <i>n</i> (%)	332 (48.5)	324 (46.2)	222 (51.2)	197 (54)
Age, mean (minimum-maximum)	7.6 (5-10)	7.9 (6-11)	8.0 (5-10)	7.7 (6-11)
SES, median (minimum-maximum)	5 (1-5)	5 (1-5)	5 (1-5)	5 (1-5)
Foreign born (child), <i>n</i> (%)	—	74 (10.5)	39 (9.0)	61 (16.7)
Foreign born (mother), <i>n</i> (%)	—	645 (92.9)	407 (94)	323 (89)
Foreign born (father), <i>n</i> (%)	—	659 (95.5)	406 (94.2)	274 (85.4)
Interethnic relationship, <i>n</i> (%)	—	40 (5.7)	33 (7.6)	74 (20.3)

Note. SES = socioeconomic status.

Table 2. Goodness-of-Fit Indices

Model	Factor Loadings and Thresholds	χ^2	<i>df</i>	<i>p</i>	CFI	RMSEA
1 ^a	Free to vary	824.24	151	.000	.931	.092
2 ^a	Equal	721.24	149	.000	.941	.085
1 ^b	Free to vary	1103.93	112	.000	.905	.092
2 ^b	Equal	1021.91	113	.000	.913	.087
1 ^c	Free to vary	459.40	49	.000	.929	.091
2 ^c	Equal	350.82	42	.000	.946	.085

Note. CFI = Comparative Fit Index; RMSEA = root mean square error of approximation.

^aComparing 4 ethnicity groups. ^bComparing boys with girls. ^cComparing Dutch with Surinamese teachers.

yields an acceptable fit for ethnicity of the child, gender of the child, and ethnicity of the teacher (restricted to Dutch and Surinamese teachers) with CFI >.900 and RMSEA ≤.091). From these results we may conclude that configural invariance of the SDQ was supported for ethnicity of the child, gender of the child, and ethnicity of teacher, which means that the five factors consisted of the same items across the different groups.

To test metric and scalar invariance, factor loadings and thresholds were constrained to be equal. Factors were allowed to correlate freely in both models. The difference χ^2 -test between Model 1 and Model 2 was significant for ethnicity of the child, $\Delta\chi^2(52) = 100.45$, $p = .000$; for gender of the child, $\Delta\chi^2(21) = 74.96$, $p = .000$; but not for ethnicity of the teacher, $\Delta\chi^2(9) = 15.62$, $p = .075$. Despite the statistical significance for ethnicity of the child and for gender of the child, we found that the fit indices show in all three cases increasing CFI values and decreasing RMSEA values, indicating that a factor loading and threshold constrained model better fits the data than an unconstrained model. From these results we may infer that a full metric model is supported by the data for ethnicity of the child, gender of the child, and ethnicity of the teacher. Because metric and scalar invariance was supported by the data, we did a final CFA over all respondents. The factor loadings and factor correlations are reported in Tables 3 and 4, respectively. The differences between the final model and the three constrained models are minor. From Table 3 it can be seen that the factor loadings are sufficiently high, mostly above .7 but with three factor loadings around .5. The factor correlations (Table 4) show that conduct problems, peer problems, and prosocial behavior are highly interrelated.

Table 3. Factor Loadings of the Strengths and Difficulties Questionnaire Scales

Emotional Symptoms	Loading	Conduct Problems	Loading	Hyperactivity/Inattention	Loading	Peer Problems	Loading	Prosocial Behavior	Loading
Somatic Worries	.56 .77	Tempers Obedient (-)	.69 .82	Restless Fidgeting	.93 .89	Alone Good friend (-)	.46 .75	Considerate Shares	.96 .84
Unhappy Clingy (-) Afraid (-)	.96 .80 .74	Fights Lies Steals	.86 .85 .77	Distracted Reflective (-) Attentive (-)	.86 .86 .82	Popular (-) Bullied Prefers adults	.98 .75 .52	Caring Kind Helpful	.83 .77 .68

Note: (-) = Reversed items

Table 4. Factor Correlations of the Strengths and Difficulties Questionnaire Scales

Factors	Correlation
Emotional Symptoms	
× Conduct Problems	.29
× Hyperactivity/Inattention	.34
× Peer Problems	.56
× Prosocial Behavior	-.24
Conduct Problems	
× Hyperactivity/Inattention	.71
× Peer Problems	.64
× Prosocial Behavior	-.78
Hyperactivity/Inattention	
× Peer Problems	.46
× Prosocial Behavior	-.50
Peer Problems	
× Prosocial Behavior	-.72

Mean Scores

Table V shows means, standard deviations, and Cronbach's alphas for the six scales of the SDQ. Alpha reliabilities were acceptable for Emotional Symptoms (.72-.78), Conduct Problems (.72-.79), and Peer Problems (.70-.75) and good for Hyperactivity/Inattention (.85-.88), Prosocial Behavior (.83-.85), and the Impact Scores (.76-.81).

Emotional symptoms were higher for Dutch than for Moroccan ($t = 2.7$; $df = 170$, $p < .01$) and Turkish ($t = 2.5$; $df = 170$; $p = .01$) children. Conduct problems were higher for Moroccan than for Dutch ($t = 5.6$; $df = 170$; $p < .001$), Turkish ($t = 6.6$; $df = 170$; $p < .001$), and Surinamese ($t = 2.8$; $df = 170$; $p < .01$) children; higher for Surinamese than for Dutch ($t = 2.2$; $df = 170$; $p = .03$) and Turkish ($t = 2.4$; $df = 170$; $p = .02$) children; and higher for boys than for girls ($t = 8.9$; $df = 170$; $p < .001$). Hyperactivity/Inattention was higher for Moroccan children than for Dutch ($t = 2.0$; $df = 170$; $p = .04$) and Turkish ($t = 2.6$; $df = 170$; $p < .01$) children and higher for boys than for girls ($t = 9.6$; $df = 170$; $p < .001$). Prosocial behavior was lower for Moroccan than for Dutch ($t = -5.1$; $df = 170$; $p < .001$), Turkish ($t = -3.2$; $df = 170$; $p < .01$), and Surinamese children ($t = -3.5$; $df = 170$; $p = .001$); lower for Turkish than for Dutch children ($t = -2.1$; $df = 170$; $p = .04$); and lower for boys than for girls ($t = -11.3$; $df = 170$; $p < .001$). Peer problems did not differ

Table 5. Means, 95% Confidence Intervals, and Cronbach's Alpha of Strengths and Difficulties Questionnaire Scales by Ethnicity and Gender

Group	Emotional		Conduct		Hyperactivity		Peer		Prosocial		Impact	
	M (95% CI)	α	M (95% CI)	α	M (95% CI)	α	M (95% CI)	α	M (95% CI)	α	M (95% CI)	α
DuTot	1.8 (1.5-2.0) _a	.77	1.1 (9-1.3) _a	.72	3.0 (2.7-3.3) _a	.85	1.5 (1.3-1.8) _a	.75	7.9 (7.6-8.3) _a	.85	5 (4-6) _a	.77
MoTot	1.4 (1.2-1.6) _b	.76	1.8 (1.6-2.1) _b	.79	3.4 (3.1-3.7) _b	.88	1.7 (1.5-2.0) _a	.71	6.9 (6.6-7.2) _b	.84	5 (3-6) _a	.76
TuTot	1.4 (1.2-1.6) _b	.72	1.0 (9-1.2) _a	.72	2.8 (2.5-3.2) _a	.87	1.6 (1.4-1.8) _a	.74	7.5 (7.1-7.8) _c	.83	3 (2-4) _b	.76
SuTot	1.5 (1.2-1.7) _{a,b}	.78	1.4 (1.2-1.6) _c	.78	3.2 (2.9-3.5) _{a,b}	.87	1.5 (1.3-1.8) _a	.70	7.7 (7.3-8.0) _{a,c}	.83	4 (3-5) _{a,b}	.81
Boys	1.4 (1.3-1.6) _a	.76	1.7 (1.6-1.9) _a	.78	3.7 (3.5-4.0) _a	.86	1.7 (1.6-1.9) _a	.71	6.9 (6.6-7.1) _a	.84	5 (5-6) _a	.77
Girls	1.6 (1.4-1.7) _a	.76	.9 (0.8-1.0) _b	.73	2.5 (2.3-2.7) _b	.86	1.5 (1.3-1.6) _b	.74	8.1 (7.9-8.4) _b	.82	3 (2-3) _b	.77
DuBoy	1.7 (1.5-2.0) _a	.76	1.5 (1.3-1.7) _a	.74	3.6 (3.3-3.9) _a	.83	1.7 (1.4-1.9) _a	.75	7.3 (6.9-7.7) _a	.83	6 (5-7) _a	.78
MoBoy	1.3 (1.1-1.6) _b	.78	2.2 (2.0-2.5) _b	.79	4.0 (3.7-4.4) _a	.88	1.9 (1.6-2.1) _a	.67	6.3 (5.9-6.6) _b	.80	6 (5-7) _a	.72
TuBoy	1.4 (1.1-1.6) _b	.67	1.4 (1.3-1.6) _c	.69	3.5 (3.1-3.8) _b	.86	1.8 (1.5-2.0) _a	.71	6.8 (6.5-7.2) _a	.83	4 (3-5) _b	.74
SuBoy	1.4 (1.1-1.7) _{a,b}	.79	1.8 (1.5-2.1) _a	.81	3.8 (3.5-4.2) _a	.87	1.7 (1.4-1.9) _a	.71	7.0 (6.6-7.4) _a	.83	5 (4-7) _a	.83
DuGirl	1.8 (1.6-2.1) _a	.79	.7 (0.5-0.8) _a	.64	2.4 (2.1-2.7) _a	.85	1.4 (1.1-1.6) _a	.74	8.6 (8.3-8.9) _a	.83	3 (2-4) _a	.74
MoGirl	1.4 (1.2-1.7) _a	.74	1.4 (1.2-1.6) _b	.75	2.8 (2.4-3.1) _a	.86	1.6 (1.4-1.8) _b	.76	7.5 (7.2-7.8) _b	.84	3 (2-4) _a	.80
TuGirl	1.5 (1.2-1.7) _a	.75	.6 (4-8) _{a,c}	.76	2.2 (1.8-2.6) _a	.87	1.5 (1.2-1.7) _b	.76	8.1 (7.8-8.5) _b	.84	2 (1-3) _a	.77
SuGirl	1.5 (1.2-1.8) _a	.77	1.0 (7-1.2) _{b,c}	.71	2.6 (2.2-2.9) _a	.85	1.4 (1.1-1.6) _{a,b}	.69	8.3 (7.9-8.7) _b	.83	3 (2-4) _a	.76

Note. Du = Dutch; Mo = Moroccan; Tu = Turkish; Su = Surinamese; Tot = total (boys and girls); Boy = boys only; Girl = girls only. Groups with similar subscripts are not significantly different.

Table 6. Relationship Between Symptom Scales and Impairment Across Ethnicity

	Emotional	Conduct	Hyperactivity	Peer	Prosocial
Dutch OR [95% CI]	1.49 [1.31, 1.68]***	1.38 [1.17, 1.64]**	1.51 [1.37, 1.68]***	1.48 [1.28, 1.69]***	1.08 [0.95, 1.23]
Moroccan OR [95% CI]	1.35 [1.19, 1.53]***	1.17 [1.02, 1.35]*	1.55 [1.40, 1.73]***	1.30 [1.13, 1.49]***	0.86 [0.76, 0.98]*
Turkish OR [95% CI]	1.32 [1.09, 1.60]**	1.01 [0.81, 1.27]	1.79 [1.52, 2.10]***	1.30 [1.09, 1.56]**	0.94 [0.79, 1.11]
Surinamese OR [95% CI]	1.59 [1.31, 1.92]***	1.32 [1.04, 1.66]*	1.64 [1.38, 1.95]***	1.42 [1.15, 1.77]**	1.18 [0.94, 1.49]

Note: * $p < .05$. ** $p < .01$. *** $p < .001$.

significantly across ethnicity but were higher for boys than for girls ($t = 3.2$; $df = 170$; $p < .01$). Impact scores were lower for Turkish than for Dutch ($t = -2.5$; $df = 170$; $p = .01$) and Moroccan children ($t = -2.2$; $df = 170$; $p = .03$) and lower for girls than for boys ($t = -5.2$; $df = 170$; $p < .001$).

Functional Equivalence: Association With Impairment

After testing for construct equivalence and comparing means, we examined whether the relationship between the different scales and the level of impairment was similar for Dutch and immigrant children. Emotional Symptoms scores, Conduct Problems scores, Hyperactivity/Inattention scores, and Peer Problems scores were positively related to impairment; odds ratios (with 95% CI in brackets) were 1.43 [1.33, 1.54], 1.24 [1.13, 1.35], 1.56 [1.46, 1.65], and 1.36 [1.25, 1.47], respectively, whereas no relationship was observed between Prosocial Behavior scores and impairment: 0.99 [0.92, 1.07]. As shown in Table 6, these relationships appeared to be similar for all ethnic groups, with the exception of a significant negative relationship between Prosocial Behavior scores and impairment across the Moroccan group and a nonsignificant positive relationship between Conduct Problems and impairment among the Turkish group.

Discussion

The aim of the present study was to examine the construct equivalence of the teacher SDQ and to study the relationship between symptom scores and impairment in an ethnically diverse sample of children residing in the Netherlands. Three main findings are of importance here: First, the equivalence of the teacher SDQ is supported across child's ethnicity and gender and also across teacher's ethnicity (i.e., Dutch and Surinamese). Second, mean scores differed across children's ethnicity and gender. Third, the relationship between the symptom scales and impairment (i.e., functional equivalence) was generally similar across ethnic groups.

Results of the multisample confirmatory factor analysis confirmed the original five-factor model of the SDQ in Dutch, Moroccan, Turkish, and Surinamese boys and girls. Therefore, it appears likely that teachers construe Emotional Symptoms, Conduct Problems, Hyperactivity/Inattention, Peer Problems, and Prosocial Behavior generally similar across children's ethnicity and gender. This is a striking finding given the low levels of assimilation of the immigrant children as indicated by the low percentage of interethnic relationships among all immigrant groups and the poor Dutch language skills among Moroccan and Turkish immigrant parents as observed in the current study. Previous studies have offered contradictory results regarding the underlying dimensions of the SDQ. The original factor structure of the SDQ was replicated among German parents (Woerner, Becker, & Rothenberger, 2004) and parents and adolescents of Dutch origin

(Muris et al., 2003) but not among Norwegian adolescents (Ronning et al., 2004), American parents (Dickey & Blumberg, 2004), Arab children, parents and teachers living in the Gaza strip (Thabet et al., 2000), and Congolese teachers (Kashala et al., 2005). These inconsistencies in findings may be due to differences in methods (e.g., confirmatory vs. exploratory analysis) or to differences in the underlying constructs. For instance, German parents and Dutch parents may construe problem behaviors similarly, whereas Congolese teachers may differ in their construction of problem behaviors. No studies exist that examined the SDQ among Moroccans, Turkish, or Surinamese children. However, our results are consistent with a study that replicated the original factor structure of the CBCL in a Turkish sample (Dumenci et al., 2004). Additionally, the present findings suggest that metric and scalar invariance was supported among native Dutch and immigrant children. This finding of invariance is in line with a study in which the metric invariance of the ADHD-IV Rating Scale was tested among Caucasian and African American students in the United States (Reid et al., 1998).

The construct equivalence also appeared to be similar for native Dutch and Surinamese immigrant teachers, suggesting that native Dutch and Surinamese immigrant teachers construe problem behaviors among Dutch, Moroccan, Turkish, and Surinamese children similarly. This finding is of particular relevance in light of research by Crijnen et al. (2000) that showed important differences in ratings of Turkish immigrant children in the Netherlands, depending on whether they were rated by Dutch or Turkish teachers. The present findings suggest that the observed differences are not likely to be due to construct nonequivalence but may be explained by ethnic differences in the perception of behavior problems (see, e.g., Mann, Ikeda, Mueller, & Takahashi, 1992).

Whereas construct equivalence was supported, mean scores differed. The cross-culturally consistent gender difference of higher scores on prosocial behavior for girls and higher scores on externalizing problems for boys are in accordance with previous findings (Crijnen, Achenbach, & Verhulst, 1997). Therefore, it seems likely that the observed differences between boys and girls reflect, at least in part, actual differences in behavior. The observation that boys score higher on impairment than girls is also consistent with this interpretation. Mean scores were neither similar across ethnicity. Moroccan children were scored relatively high on externalizing problems and low on prosocial behavior, whereas Dutch children were rated relatively high on internalizing problems and prosocial behavior. Previous studies on problem behavior among children of different ethnic origin in the Netherlands have been inconclusive (Bengi-Arslan et al., 1997; Crijnen et al., 2000; Stevens et al., 2003). For example, teacher reports were similar for Turkish immigrant and Dutch native children (Crijnen et al., 2000), whereas CBCL scores were higher for Turkish immigrant than for Dutch native children. Likewise, teacher ratings were higher for Moroccan immigrant youth than for Turkish immigrant youth and Dutch native youth, whereas self-reports were lower for Moroccan immigrant youth (Stevens et al., 2003). Therefore, it can be questioned whether the reported ethnic differences in problem behavior reflect actual differences in behavior, especially given recent findings that the prevalence of internalizing and externalizing problems does not differ according to best-estimate diagnoses (Zwirs et al., 2007). This may suggest that the observed differences across ethnicity can be explained at least to some degree by a rater bias (see, e.g., Sonuga-Barke & Minocha, 1993). However, the lack of observational data is an important limitation of this study and precludes any firm conclusions about whether observed differences are due to a rater bias or actual differences in behavior. Therefore, further research that includes observational data is needed. Moreover, caution is further warranted when interpreting these differences in scores, because they can also reflect method bias or item bias rather than valid cross-cultural or cross-gender differences (van de Vijver & Leung, 2001). A method bias may imply, for instance, that teachers may use different response styles (i.e., acquiescence, social desirable responding, and extreme response bias) across different

ethnic and or gender groups. An item bias refers to inconsistencies at the item level, such as the unsuitability of an item in a particular culture. However, because the majority of the teachers were of native origin, the risk of method or item bias seems limited. Still, future studies that examine response sets and differential item functioning will be important in ruling out alternative interpretations of score differences across different groups (van de Vijver & Leung, 2001).

Three other important limitations should be kept in mind when considering the findings. First, because the number of Moroccan and Turkish teachers was too small, construct equivalence could only be tested in Dutch and Surinamese teachers. However, because teachers in the Netherlands are largely of Dutch origin and play an important role in the detection of behavior problems, their construction of behavior problems is of great importance. Nevertheless, additional research is needed to determine whether the equivalence of the factor structure is supported among Turkish and Moroccan teachers. Given that cross-ethnic differences have been observed not only by teachers (Crijnen et al., 2000; Stevens et al., 2003; Zwirs, Burger, Schulpen, & Buitelaar, 2006), but also by parents (Bengi-Arslan et al., 1997; Stevens et al., 2003; Zwirs, Burger, Buitelaar, & Schulpen, 2006) and have been shown in self-reports (Stevens et al., 2003), future studies should also determine whether construct equivalence is confirmed among parents and youth of Dutch, Moroccan, Turkish and Surinamese background.

Second, because we sampled from inner-city low SES neighborhoods, the results cannot be generalized to other neighborhoods with middle and higher SES levels without further study. However, because most immigrants in the Netherlands live in low SES urban areas, generalizing to middle and higher SES is of limited importance. Moreover, the strength of sampling from low SES neighborhoods only is that SES is similar across ethnic groups, enabling us to test ethnic differences in relative separation from SES. As has been noted by Epstein et al. (2005), SES and ethnicity are highly correlated in most studies, making it difficult to control for SES.

Third, despite the rather rudimentary manner of estimating SES by postal codes of the schools, the SES indicated by the school is likely to closely resemble that of the household given that most children in the Netherlands attend a school in their own neighborhood. It is unlikely that children from higher SES neighborhoods will attend a school in a lower SES neighborhood. Still, the present findings should be replicated in other studies that can take individual measures of SES into account.

The findings of the present study have some clinical implications that are not restricted to the Netherlands but apply also to the United States and other multiethnic societies, given that in most countries the majority of schoolteachers are of native origin. The current findings yield no support for the idea that gender and ethnic differences in externalizing and internalizing problems are explained by construct nonequivalence of the SDQ across gender or ethnicity. With respect to the observed gender differences, it seems likely that these differences reflect actual differences in behavior between boys and girls, because boys were also higher rated on impairment problems and because these findings of higher problem scores for boys are in line with previous studies using different methods. However, it can be questioned whether the ethnic differences in mean scores reflect actual differences in behavior, given that previous studies have yielded contradictory results depending on which informant was used and that recent findings based on best-estimate diagnoses suggest the prevalence of internalizing and externalizing problems to be similar across the respective ethnic groups. As a result, the possibility of a rater bias across ethnicity becomes more likely, implying that for instance stereotypes about particular ethnic groups may explain differences in behavior scores. Therefore, correction for mean differences through separate norms across ethnicity may be warranted, because ratings may be inflated for children of particular ethnic groups. In addition, because the relationship between symptoms and impairment may vary across ethnicity according to teacher reports as indicated by the present results, another clinical implication is that the assessment of functional impairment is of particular

importance when screening problem behavior among children of different ethnic origin. Still, we caution that these results should be replicated in studies using vignettes or observational data.

Declaration of Conflicting Interests

Dr Buitelaar has served as consultant and has been on advisory boards for Eli Lilly, Janssen CilagBV, UCB, and Shire in the past two years. The other authors have no financial relationships to disclose.

Financial Disclosure/Funding

This study was financially supported by the Netherlands Organization for Scientific Research (ZON-MW), grant number 99-9.1-64, by the Foundation for Children's Welfare Stamps Netherlands and by the Province of Utrecht, the Netherlands.

References

- Achenbach, T. M. & Rescorla, L. A. (2001). *Manual for the ASEBA preschool forms & profiles*. Burlington, University of Vermont, Research Center for Children, Youth, & families.
- Beiser, M., Dion, R., & Gotowiec, A. (2000). The structure of attention-deficit and hyperactivity symptoms among native and non-native elementary school children. *Journal of Abnormal Child Psychology*, 28, 425-437.
- Bengi-Arslan, L., Verhulst, F. C., Van der Ende, J., & Erol, N. (1997). Understanding childhood (problem) behaviors from a cultural perspective: comparison of problem behaviors and competencies in Turkish immigrant, Turkish and Dutch children. *Social Psychiatry Psychiatric Epidemiology*, 32, 477-484.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238-246.
- Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R. (2002). *Cross-cultural psychology; Research and applications*. Cambridge, U.K.: Cambridge University Press.
- Bird, H. R. (1996). Epidemiology of childhood disorders in a cross-cultural context. *Journal of Child Psychology and Psychiatry*, 37, 35-49.
- Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic applications and programs*. Hillsdale, NJ: Lawrence Erlbaum.
- Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods & Research*, 36, 462-494.
- Crijnen, A. A. M., Achenbach, T. M., & Verhulst, F. C. (1997). Comparisons of problems reported by parents of children in 12 cultures: Total problems, externalizing and internalizing. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36, 1269-1277.
- Crijnen, A. A., Achenbach, T. M., & Verhulst, F. C. (1999). Problems reported by parents of children in multiple cultures: The Child Behavior Checklist syndrome constructs. *American Journal of Psychiatry*, 156, 569-574.
- Crijnen, A. A., Bengi-Arslan, L., & Verhulst, F. C. (2000). Teacher-reported problem behaviour in Turkish immigrant and Dutch children: a cross-cultural comparison. *Acta Psychiatrica Scandinavica*, 102, 439-444.
- 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 and Adolescent Psychiatry*, 43, 1159-1167.
- Dumenci, L., Erol, N., Achenbach, T. M., & Simsek, Z. (2004). Measurement structure of the Turkish translation of the Child Behavior Checklist using confirmatory factor analytic approaches to validation of syndromal constructs. *Journal of Abnormal Child Psychology*, 32, 335-340.
- Epstein, J. N., March, J. S., Conners, C. K., & Jackson, D. L. (1998). Racial differences on the Conners Teacher Rating Scale. *Journal of Abnormal Child Psychology*, 26, 109-118.
- Epstein, J. N., Willoughby, M., Valencia, E. Y., Tonev, S. T., Abikoff, H. B., Arnold, L. E., & Hinshaw, S. P. (2005). The role of children's ethnicity in the relationship between teacher ratings of

- attention-deficit/hyperactivity disorder and observed classroom behavior. *Journal of Consulting and Clinical Psychology*, 73, 424-434.
- Goodman, R. (1994). A modified version of the Rutter parent questionnaire including extra items on children's strengths: A research note. *Journal of Child Psychology and Psychiatry*, 35, 1483-1494.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38, 581-586.
- Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 1337-1345.
- Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *British Journal of Psychiatry*, 177, 534-539.
- Goodman, R., Renfrew, D., & Mullick, M. (2000). Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. *European Child and Adolescent Psychiatry*, 9, 129-134.
- Gross, D., Fogg, L., Young, M., Ridge, A., Cowell, J. M., Richardson, R., & Sivan, A. (2006). The equivalence of the Child Behavior Checklist/11/2-5 across parent race/ethnicity, income level, and language. *Psychological Assessment*, 18, 313-323.
- Hartman, C. A., Hox, J., Auerbach, J., Erol, N., Fonseca, A. C., Mellenbergh, G. J., . . . Sergeant, J. A. (1999). Syndrome dimensions of the Child Behavior Checklist and the Teacher Report Form: a critical empirical evaluation. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 40, 1095-1116.
- Ivanova, M. Y., Achenbach, T. M., Dumenci, L., Almqvist, F., Weintraub, S., Bilenberg, N., . . . Verhulst, F. C. (2007). Testing the 8-syndrome structure of the child behavior checklist in 30 societies. *Journal of Clinical Child and Adolescent Psychology*, 36, 405-417.
- Joreskog, K. G., & Sorbom, D. (1993). *LISREL 8 user's reference guide*. Chicago, IL: Scientific Software International.
- Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Thousand Oaks, CA: Sage.
- Kashala, E., Elgen, I., Sommerfelt, K., & Tylleskar, T. (2005). Teacher ratings of mental health among school children in Kinshasa, Democratic Republic of Congo. *European Child and Adolescent Psychiatry*, 14, 208-215.
- Kataoka, S. H., Zhang, L., & Wells, K. B. (2002). Unmet need for mental health care among U.S. children: variation by ethnicity and insurance status. *American Journal of Psychiatry*, 159, 1548-1555.
- Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York, NY: Guilford Press.
- Knol, F. A. (1998). *Van hoog naar laag; van laag naar hoog (From high to low; from low to high)*. The Hague, Netherlands: Sociaal en Cultureel Planbureau.
- Kvernmo, S., & Heyerdahl, S. (1998). Influence of ethnic factors on behavior problems in indigenous Sami and majority Norwegian adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, 37, 743-751.
- Mann, E. M., Ikeda, Y., Mueller, C. W., & Takahashi, A. (1992). Cross-cultural differences in rating hyperactive-disruptive behaviors in children. *American Journal of Psychiatry*, 149, 1539-1542.
- Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Benlert's (1999) findings. *Structural Equation Modeling*, 11, 320-341.
- Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, 58, 525-543.
- Murad, S. D., Joung, I. M., van Lenthe, F. J., Bengi-Arslan, L., & Crijnen, A. A. (2003). Predictors of self-reported problem behaviours in Turkish immigrant and Dutch adolescents in the Netherlands. *Journal of Child Psychology and Psychiatry*, 44, 412-423.

- Muris, P., Meesters, C., & Van Den, B. F. (2003). The Strengths and Difficulties Questionnaire (SDQ)—Further evidence for its reliability and validity in a community sample of Dutch children and adolescents. *European Child and Adolescent Psychiatry*, 12, 1-8.
- Muthén, L. K., & Muthén, B. O. (1998-2006). *Mplus user's guide* (4th ed.). Los Angeles, CA: Muthén & Muthén.
- Olfson, M., Gameroff, M. J., Marcus, S. C., & Jensen, P. S. (2003). National trends in the treatment of attention deficit hyperactivity disorder. *American Journal of Psychiatry*, 160, 1071-1077.
- Reid, R., Casat, C. D., Norton, H. J., Anastopoulos, A. D., & Temple, E. P. (2001). Using behavior rating scales for ADHD across ethnic groups: The IOWA Conners. *Journal of Emotional and Behavioral Disorders*, 9, 210-219.
- Reid, R., DuPaul, G. J., Power, T. J., Anastopoulos, A. D., Rogers-Adkinson, D., Noll, M. B., & Riccio, C. (1998). Assessing culturally different students for attention deficit hyperactivity disorder using behavior rating scales. *Journal of Abnormal Child Psychology*, 26, 187-198.
- Roberts, R. E., Attkisson, C. C., & Rosenblatt, A. (1998). Prevalence of psychopathology among children and adolescents. *American Journal of Psychiatry*, 155, 715-725.
- Ronning, J. A., Handegaard, B. H., Sourander, A., & Mørch, W. T. (2004). The Strengths and Difficulties Self-Report Questionnaire as a screening instrument in Norwegian community samples. *European Child and Adolescent Psychiatry*, 13, 73-82.
- Rowland, A. S., Umbach, D. M., Stallone, L., Naftel, A. J., Bohlig, E. M., & Sandler, D. P. (2002). Prevalence of medication treatment for attention deficit-hyperactivity disorder among elementary school children in Johnston County, North Carolina. *American Journal of Public Health*, 92, 231-234.
- Safer, D. J., & Malever, M. (2000). Stimulant treatment in Maryland public schools. *Pediatrics*, 106, 533-539.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8, 23-74.
- Sonuga-Barke, E. J. S., & Minocha, K. (1993). Inter-ethnic bias in teachers' ratings of childhood hyperactivity. *British Journal of Developmental Psychology*, 11, 187-200.
- Steenkamp, J-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25, 78-90.
- Stevens, G. W., Pels, T., Bengi-Arslan, L., Verhulst, F. C., Vollebergh, W. A., & Crijnen, A. A. (2003). Parent, teacher and self-reported problem behavior in the Netherlands: Comparing Moroccan immigrant with Dutch and with Turkish immigrant children and adolescents. *Social Psychiatry and Psychiatric Epidemiology*, 38, 576-585.
- Thabet, A. A., Stretch, D., & Vostanis, P. (2000). Child mental health problems in Arab children: application of the Strengths and Difficulties Questionnaire. *International Journal of Social Psychiatry*, 46, 266-281.
- van de Vijver, F. J. R., & Leung, K. (2001). Personality in cultural context: Methodological issues. *Journal of Personality*, 69, 1007-1031.
- Van Widenfelt, B. M., Goedhart, A. W., Treffers, P. D., & Goodman, R. (2003). Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *European Child and Adolescent Psychiatry*, 12, 281-289.
- Verhulst, F. C., Achenbach, T. M., van der, E. J., Erol, N., Lambert, M. C., Leung, P. W., Zubrick, S. R. (2003). Comparisons of problems reported by youths from seven countries. *American Journal of Psychiatry*, 160, 1479-1485.
- Weisz, J. R., Chaiyasit, W., Weiss, B., Eastman, K. L., & Jackson, E. W. (1995). A multimethod study of problem behavior among Thai and American children in school: Teacher reports versus direct observations. *Child Development*, 66, 402-415.
- Woerner, W., Becker, A., & Rothenberger, A. (2004). Normative data and scale properties of the German parent SDQ. *European Child and Adolescent Psychiatry*, 13(Suppl. 2), II3-II10.

- Woerner, W., Fleitlich-Bilyk, B., Martinussen, R., Fletcher, J., Cucchiaro, G., Dalgarrondo, P., . . . (2004). The Strengths and Difficulties Questionnaire overseas: Evaluations and applications of the SDQ beyond Europe. *European Child and Adolescent Psychiatry*, 13(Suppl. 2), II47-II54.
- Zwirs, B. W. C., Burger, H., Buitelaar, J. K., & Schulp, T. W. J. (2006). Ethnic differences in parental detection of externalizing disorders. *European Child & Adolescent Psychiatry*, 15, 418-426.
- Zwirs, B. W. C., Burger, H., Schulp, T. W. J., & Buitelaar, J. K. (2006). Different treatment thresholds in non-western children with behavioral problems. *Journal of the American Academy of Child and Adolescent Psychiatry*, 45, 476-483.
- Zwirs, B., Burger, H., Schulp, T., & Buitelaar, J. (2008). Developing a brief cross-culturally validated screening tool for externalizing disorders in children. *Journal of the American Academy of Child and Adolescent Psychiatry*, 47, 309-316.
- Zwirs, B. W., Burger, H., Schulp, T. W., Wiznitzer, M., Fedder, H., & Buitelaar, J. K. (2007). Prevalence of psychiatric disorders among children of different ethnic origin. *Journal of Abnormal Child Psychology*, 35, 556-566.

Bios

Barbara Zwirs is a developmental psychologist and cultural anthropologist who received her PhD in child psychiatry from the University of Utrecht, the Netherlands, in 2006. Since 2006 she has been an assistant professor in criminology at Erasmus University Rotterdam, the Netherlands. Her research interests concern the intersection of criminology, developmental psychology, and social psychology.

H. Burger received his PhD in medicine from Erasmus University, Rotterdam, the Netherlands, in 1995. Between 1998 and 2000 he worked predominantly as a research consultant, collaborating with a variety of researchers at the University of Utrecht, the Netherlands. At present, he is an associate professor at Groningen State University, the Netherlands. His main interest is in psychiatric epidemiology, in particular the study of effects of psychosocial stressors, and in methods of research.

Tom Schulp, MD, MPH, PhD, was professor of pediatrics at Utrecht University with main interest in culture and health. His Center for Migration and Child Health conducted several studies about multicultural aspects of health and disease of children in the Netherlands. Three years ago, he retired from active clinical and scientific work.

Ad Vermulst received his PhD in the social sciences from the University of Groningen, the Netherlands. His main interest is in statistical methodology (latent variable modeling, longitudinal statistical modeling). He is a senior researcher of the Behavioural Science Institute of the University of Nijmegen and is publishing about topics like personality, family stress, child rearing, and youth risk behavior.

R. A. HiraSing is pediatrician, youth health doctor, and professor of Youth Health Care, VU University Medical Centre, Amsterdam. He is head of the Youth Health Department. His main research interests are overweight and obesity, type 2 diabetes, and psychosocial problems, all in children with focus on prevention and evidence based guidelines.

Jan Buitelaar, MD, PhD, is a child and adolescent psychiatrist with a longstanding clinical and research interest in neuropsychiatric disorders as attention-deficit/hyperactivity disorder and autism.