



Published in final edited form as:

Psychol Assess. 2007 June ; 19(2): 189–198.

Examining the Structural Validity of the Strengths and Difficulties Questionnaire (SDQ) in a U.S. Sample of Custodial Grandmothers

Patrick A. Palmieri and

Summa-Kent State Center for the Treatment and Study of Traumatic Stress, Summa Health System, Akron, Ohio; and Department of Psychology, Kent State University

Gregory C. Smith

College of Education, Health, and Human Services, Kent State University.

Abstract

The authors examined the structural validity of the parent informant version of the Strengths and Difficulties Questionnaire (SDQ) with a sample of 733 custodial grandparents. Three models of the SDQ's factor structure were evaluated with confirmatory factor analysis based on the item covariance matrix. Although indices of fit were good across all 3 models, a model that included a newly hypothesized positive construal method factor in addition to the 4 symptom factors (Emotional Symptoms, Conduct Problems, Hyperactivity-Inattention, Peer Problems) and the single Prosocial Behavior factor originally intended for the SDQ provided the best representation of this instrument's latent structure. Structural validity was further evidenced by measurement invariance across grandparent race and grandchild gender and age, a conceptually meaningful pattern of cross-scale correlations, and the acceptable internal reliability estimates found for each subscale. Measurement and clinical implications of the results are discussed.

Keywords

SDQ; custodial grandparents; CFA; psychometrics; factor structure

The Strengths and Difficulties Questionnaire (SDQ) is a promising new instrument for assessing the psychological adjustment of children and adolescents, first published in 1997 by British psychiatrist Robert Goodman. The aim of the current study was to examine the structural validity of the parent informant version of the SDQ with a large national sample of custodial grandmothers providing full-time care to a grandchild in the absence of the child's biological parents. Discrepancies in the literature regarding the structural validity of the SDQ and recent growth in the population of grandparents as caregivers underscore the need for this research. Confirmatory factor analysis (CFA) was used to test the purported factor structure of the SDQ, and the cross-scale correlations between subscales and the internal reliabilities of all subscales were also examined to evaluate this instrument's structural validity.

A Brief Description of the SDQ

The SDQ may be administered to parents and teachers of 4- to 16-year-olds and to 11- to 16-year-olds themselves. It contains 25 items, selected on the basis of both contemporary diagnostic criteria and factor analysis, divided equally among five scales such that subscale scores are generated for Emotional Symptoms, Conduct Problems, Hyperactivity-Inattention,

Peer Problems, and Prosocial Behavior. Ten items are worded to reflect strengths of the child (with 5 being reverse-scored as problems), 14 reflect difficulties, and 1 is neutral but scored as a difficulty on the Peer Problems subscale. The inclusion of positively worded items was done to emphasize desirable traits rather than to focus solely on deficits, thereby increasing the acceptability of the SDQ to parents and other informants (Goodman, 1999). An extended SDQ also exists, which assesses the impact of symptoms on social and educational function, distress, and burden on others (Goodman, 1999).

The SDQ's popularity has soared, as indicated by the fact that it (a) exists in 40 languages, (b) has normative data from diverse countries including the United States, (c) is available free of charge from the Internet (see <http://www.sdqinfo.com>) for noncommercial use; and (d) is being used in the National Health Interview Survey (NHIS; National Center for Health Statistics, 2003). This instrument offers several advantages over conceptually similar yet more established measures such as the Rutter (1967) and Achenbach (1991) questionnaires. These include a more balanced focus on strengths as well as difficulties; better coverage of inattention, peer problems, and prosocial behavior; a shorter, more acceptable format focusing on positive as well as negative child attributes; and a single form for parents and teachers to increase parent-teacher concurrence (Goodman, Meltzer, & Bailey, 2003). The SDQ's brevity and coverage of strengths and difficulties make it well suited for conducting epidemiological research and for screening low-risk children in the general population, in which the majority of children are healthy.

Numerous studies from diverse countries have yielded favorable results regarding the SDQ's construct validity and clinical utility (for international reviews, see Marzocchi et al., 2004; Obel et al., 2004; Woerner, Fleitlich-Bilyk, et al., 2004). It has been shown to correlate substantially with more established indices of childhood psychopathology such as the Rutter (1967) and Achenbach (1991) questionnaires (see Goodman, 1997, 1999), to discriminate well between children with and without psychopathology (Goodman, 1997, 2001; Goodman et al., 2003), to be effective in screening for disorders in community samples (Goodman, Ford, Simmons, Gatward, & Meltzer, 2000), and to demonstrate sensitivity as a clinical outcome measure (Mathai, Anderson, & Bourne, 2003). Use of the SDQ subscales as outcome measures, as well as any other use that implies distinct factors, is nonetheless suspect until the structural validity of the SDQ has been established.

Past Studies of the SDQ's Structural Validity

In contrast to the strong evidence for the SDQ's clinical utility, studies of its structural validity have yielded mixed results. Whereas the magnitude and direction of the cross-scale correlations among the SDQ's subscales have generally been found to be (a) conceptually meaningful, (b) consistent with current knowledge of comorbidity, and (c) indicative of distinct constructs (Goodman, 2001; Hawes & Dadds, 2004; Muris, Meesters, & van den Berg, 2003; van Widenfelt, Goedhart, Treffers, & Goodman, 2003), low internal consistency coefficients have been observed for the parent and self-report Conduct Problems subscale and for the self-report Peer Problems subscale (Goodman, 2001; Koskelainen, Sourander, & Kaljonen, 2000; Koskelainen, Sourander & Vauras, 2001; Malmberg, Rydell, & Smedje, 2003; Muris, Meesters, Eijkelenboom, & Vincken, 2004; Muris et al., 2003; Smedje, Broman, Hetta, & von Knorring, 1999; van Widenfelt et al., 2003). Although these low internal consistency values may be due to the scant items on each SDQ subscale, it is also possible that these subscales measure more heterogeneous content than intended (Smedje et al., 1999; van Widenfelt et al., 2003). It has also been argued that these low reliability coefficients may be due to several positively worded reverse-scored items located on the Conduct Problems and Peer Problems subscales (Muris et al., 2004). Thus, the structural validity of the SDQ is suspect with respect to the troublesome internal consistency findings reported for these two subscales.

Studies examining the SDQ's proposed factor structure have likewise yielded doubts regarding this instrument's structural validity. Although some investigators using a forced five-factor solution and principal components analysis (PCA) have essentially confirmed Goodman's (1997) predicted five-factor structure with minimal cross-loadings observed among subscales (Becker, Woerner, Hasselhorn, Banaschewski, & Rothenberger, 2004; Goodman, 2001; Hawes & Dadds, 2004; Koskelainen et al., 2001; Muris et al., 2003; Smedje et al., 1999), less satisfactory results have emerged from studies in which the number of factors was unspecified. Woerner, Becker, and Rothenberger (2004) reported findings consistent with the proposed five-factor structure with their use of the German parent informant version of the SDQ, which is based on a community sample of 930 children and adolescents between 6 and 16 years of age. However, in their study of the self-report version of the SDQ among Finnish 13- to 17-year-olds, Koskelainen et al. (2001) found a three-factor solution (i.e., mixed Hyperactivity-Conduct, Prosocial, and mixed Emotional Symptoms-Peer Problems) that was similar for boys and girls only when the number of factors was unspecified. In addition, Muris et al. (2004) found a four-factor solution (i.e., Emotional Symptoms; Prosocial Behavior, including reversed items from other scales; Hyperactivity-Inattention; and mixed Peer Problems-Conduct Problems) to be more satisfactory than the predicted five-factor solution for the self-report SDQ among a large sample of nonclinical children ages 8 to 13 years old in the Netherlands.

In the only published study of the SDQ within the United States, Dickey and Blumberg (2004) examined the factor structure of the SDQ parent version among a large representative sample drawn from the 2001 NHIS. After conducting a three-step analytic procedure that respectively included PCA, exploratory factor analysis (EFA), and eventually CFA, they concluded that parental informants are likely to report on three separate but correlated dimensions: externalizing problems (Hyperactivity-Inattention and Conduct Problems), internalizing problems (Emotional Problems and Peer Problems), and a positive construal method factor. However, Dickey and Blumberg acknowledged that their failure to replicate the predicted five-factor solution observed in British, German, and Swedish samples might be due to the fact that they used the American English version of the SDQ, in which several items from the original British English version were modified to be more understandable to American parents and indicative of behaviors among children and youth in the United States. In their words, "the component scales published and validated in Britain may not be entirely appropriate for a sample of American children" (p. 1166).

In summary, studies examining both the internal consistency reliability and proposed factor structure of the SDQ have raised serious questions concerning its structural validity while singling out difficulties that can arise regarding the positively worded items. Even Goodman (2001) acknowledged that raters vary in their readiness to attribute positive qualities such that the Prosocial factor also functions as a Positive Construal factor, with the latter also including substantial loadings for positively worded items intended for other subscales. Dickey and Blumberg (2004) similarly concluded that "because answers to 8 of the 10 positively worded items were most strongly associated with this factor, the likelihood that this factor represents a methodological artifact is increasingly strong" (p. 1165). Thus, although the SDQ's emphasis on positive attributes was intended to increase its acceptability to respondents, recent evidence casts doubt on the utility of this feature.

The Present Study

Although prior studies have involved diverse informant groups (children, parents, and teachers) from various countries, there has been no published research on the SDQ's psychometric properties with surrogate parents as informants. Examining the SDQ's structural validity with parental surrogates is critical in view of Goodman's (2001) findings, which suggested that the SDQ factor structure may differ considerably by the type of informant (e.g., parents vs. teachers

vs. youth). This concern is especially relevant given that SDQ data and other health-related information about the randomly selected children in the 2001 NHIS were obtained from grandparent informants in 4.4% of the nationally representative sample (Bourdon, Goodman, Rae, Simpson, & Koretz, 2005). Furthermore, well-designed scales assessing clearly defined constructs are expected to produce factor structures that are invariant across different populations (Gorsuch, 1983). Thus, the use of a large national sample of custodial grandmothers as nontraditional informants is an important aspect of the present study.

Custodial grandparents, who provide full-time care of a grandchild in the absence of that child's parents, are also known as "skipped generation" grandparents (Pebley & Rudkin, 1999). Not only has the number of these grandparents risen dramatically over the past few decades (U.S. Bureau of the Census, 2004), there is also evidence that the grandchildren in their care are at substantial risk for psychological problems arising from parental dysfunction (Bratton, Ray, & Moffit, 1998; Brown-Standridge & Floyd, 2000; Ghuman, Weist, & Shafer, 1999). These risks include exposure to prenatal toxins, traumatic early childhood experiences, little or no appropriate interaction with parents, family conflicts, uncertainty about their future, and societal stigma (Hayslip, Shore, Henderson, & Lambert, 1998; Hirshorn, 1998; Smith, Savage-Stevens, & Fabian, 2002). Thus, a need exists to determine whether measures of children's psychological adjustment such as the SDQ may be meaningfully applied to this unique and rapidly growing target population.

Another key feature of this study is that in contrast to most of the past research on the SDQ's factor structure, a rigorous CFA approach was used instead of traditional EFA. Unlike EFA, which is used primarily for reducing the number of scaled items, CFA assumes that the number of factors in a model is hypothesized a priori, along with specific expectations about which variables will load onto which factors. This distinction is important because structural equation modeling (SEM)-based CFA permits a rigorous test of whether indicator variables load onto latent constructs (factors) exactly as predicted in the absence of measurement error (Bollen, 1989; Gorsuch, 1983). Unlike EFA, SEM also offers corrections for nonnormally distributed data (Bentler & Dudgeon, 1996), a likely problem among mental health measures like the SDQ (Achenbach, Howell, Quay, & Conners, 1991) that might otherwise yield inflated chi-squares and deflated standard errors, thereby increasing Type I error, and imprecise factor loadings (Schumacker & Beyerlein, 2000).

We tested three models, shown in Figure 1, which were suggested by both the extant literature and past empirical research on the SDQ's factor structure. Model 1 is a five-factor, higher order solution, which corresponds to Goodman's (1997) claim that the SDQ's four problem-oriented scales represent behavioral difficulties, whereas the Prosocial subscale is presumed to exist as a conceptually distinct construct representing behavioral strengths. Model 1 thus contains a hypothesized second-order factor labeled Difficulties, as well as an independent first-order factor labeled Strengths, which represents the Prosocial subscale. The Strengths factor is not incorporated in the reverse direction into the second-order Difficulties factor given Goodman's (1997) claim that "the absence of prosocial behaviors is conceptually different from the presence of psychological difficulties" (p. 582). Instead, the first-order Strengths factor is hypothesized to covary with the second-order Difficulties factor. Note, however, that this structure is statistically equivalent to a model with five first-order factors subsumed by a second-order factor.

Model 2 is a lower order version of Model 1, in which the relationships among the five problem-oriented factors are explained by their intercorrelations rather than by an overarching second-order factor. This slightly less parsimonious model emphasizes the potential importance of the distinct symptom dimensions rather than of a general Difficulties factor, and in our view, it is the best representation of the model that has been supported in several previous EFA and PCA

studies (e.g., Goodman, 2001; Hawes & Dadds, 2004; Koskelainen et al., 2001; Muris et al., 2003; Smedje et al., 1999). Although Goodman's (1997) claims regarding the conceptual framework of the SDQ are more consistent with Model 1, studies by Goodman (2001) as well as others are more consistent with Model 2.

Model 3 is tested in light of the questions regarding the utility of the SDQ Prosocial subscale that emerged from prior research. This six-factor model includes all five correlated factors but also specifies an uncorrelated method factor on which all positively worded items load (i.e., five Prosocial items and five reverse-scored items from the Problem-Oriented factors). Thus, all items load on one substantive factor, and the positively worded items also load on the proposed method factor. Model 3 posits that these positively worded items might reflect method variance rather than conceptually distinct dimensions. This model is sensible given Dickey & Blumberg's (2004) findings that 8 of the 10 positively worded SDQ items were most strongly associated with the positive construal factor, suggesting that a positive construal methodological factor may exist, as suspected by Goodman (1997).

We determined that testing these three models with CFA would help establish whether the purported Prosocial factor is a meaningful component of the SDQ or is instead better interpreted as a methodological artifact. In so doing, we would also determine which model provides the best representation of the underlying structure of the SDQ. After determining the best-fitting model for the entire sample, we also sought to determine whether measurement invariance exists with respect to grandchildren's gender and age and grandparent's race.

Method

Sample

The participants were a national sample of 733 grandmothers (M age = 56.1 years, SD = 8.1) providing full-time care to a grandchild in the absence of that grandchild's parents for at least 3 months. They were recruited for a National Institute of Mental Health-funded study of stress and coping among custodial grandparents that used a combination of convenience-based (e.g., social service agencies; Internet, radio, TV, and newspaper ads) and probability-based (random recruitment letters) sampling methods.

A descriptive summary of participating grandmothers and the target grandchildren is presented in Table 1. By study design, half of the grandmothers were Black, and half were White. The sample was from 48 states and was diverse in terms of residential locale (urban: 47.8%, suburban: 19.2%, rural: 32.5%) and marital status (e.g., married: 48.0%, widowed: 13.9%, divorced: 21.7%). If a grandmother was caring for more than one grandchild, then the target grandchild was selected with the most recent birthday technique (Kish, 1965). The target grandchildren were 391 girls and 342 boys (M age = 9.8 years, SD = 3.7, range = 4-16 years), and length of care ranged from 3 months to 16 years (M = 6.4 years, SD = 4.0). Reasons for care varied, and most grandmothers gave multiple reasons for providing full-time care to the target grandchild. The respondents were caregivers largely as a result of predicaments of the parent generation (e.g., substance abuse: 55.4%, incarceration: 42.6%).

Procedure

Grandmothers completed the original English-language parent informant version of the SDQ as part of a larger telephone interview with professionally trained interviewers at a major public research university in northeastern Ohio. A description of the study was read before the interview began, and oral informed consent was obtained. As noted earlier, the SDQ contains 25 items, divided among five subscales each. Respondents rated each item with respect to the target grandchild (e.g., *Often unhappy, downhearted or tearful*) on a 3-point scale ranging from

0 (*not true*) to 2 (*certainly true*). The scoring procedures, as well as all 25 items from the English-language version of the SDQ, are available at www.sdqinfo.com. These scoring procedures were used for the present analysis.

Data Analysis

To evaluate all three models, we submitted SDQ item covariances to LISREL software (Version 8.71; Jöreskog & Sörbom, 1993) for CFA with the use of robust maximum-likelihood estimation. Each item was specified to load on only one symptom factor; symptom factors were allowed to correlate with each other but not with the method factor (when present); and error covariances were constrained to zero.

Model fit was assessed with several fit indices, including chi-square, Satorra-Bentler (S-B) chi-square (which corrects for nonnormal data; Satorra & Bentler, 1988), root-mean-square error of approximation (RMSEA; Steiger, 1990), standardized root-mean-square residual (SRMR; Bentler, 1990), comparative fit index (CFI; Bentler, 1990), nonnormed fit index (NNFI; Bentler & Bonett, 1980), and Akaike information criterion (AIC; Akaike, 1987). RMSEA values of .08 or lower indicate adequate fit; values of .05 or lower indicate excellent fit (Browne & Cudeck, 1993). Hu and Bentler (1999) suggested that SRMR values of .08 or lower and CFI and NNFI values of .95 or higher indicate good fit, though other researchers have suggested cutoffs of .05 for SRMR and .90 for CFI and NNFI (e.g., Jöreskog, Sörbom, du Toit, & du Toit, 2000). For the AIC, which is useful for comparing nested or nonnested models, smaller values indicate better fit. Adequately fitting nested models were also compared with the use of scaled S-B chi-square difference tests (Satorra & Bentler, 2001). After determining the best-fitting model, we assessed measurement invariance for grandparent race (Black and White), grandchild gender, and grandchild age group (4-7, 8-10, 11-14, and 15-17). These age groupings are the same as those used to report on the SDQ normative data, available at www.sdqinfo.com.

To further examine the structural validity of the SDQ, we also examined the internal consistency reliabilities for each subscale, as well as the pattern of cross-scale correlations. The latter included both error-free factor correlations derived from LISREL, as well as Pearson product-moment correlation coefficients.

Results

SDQ item-level means, standard deviations, skewness, and kurtosis are reported in Table 2 for the entire sample. The data were nonnormally distributed (skewness and kurtosis tests of univariate normality were significant at $p < .001$ for all items). Fit statistics for all models are presented in Table 3. Models 1 and 2 both provided adequate model-data fit according to the RMSEA, SRMR, CFI, and NNFI, but Model 2 was deemed superior on the basis of the chi-square difference test, scaled S-B $\chi^2_{\text{diff}}(5, N = 733) = 59.69, p < .001$, and a lower AIC value (1029.71 vs. 1077.84). Separate model fit indices by grandchildren's gender and age and grandmothers' race also suggested the superiority of Model 2. Thus, model fit was slightly better when covariance was assumed in the four symptom subscales than when these four subscales were predicted to load onto a second-order Difficulties factor. As seen in Table 4, all items loaded significantly on their respective factors in Model 2: Emotional Symptoms averaged .59, Conduct Problems averaged .60, Hyperactivity-Inattention averaged .69, Peer Problems averaged .50, and Prosocial Behavior averaged .57. Modification indices suggested that model fit would not improve substantially if items were allowed to load on different or multiple factors.

Model 3, which encompasses the potential positive construal method factor, enjoyed greater empirical support than did Models 1 and 2. Not only were the RMSEA (.046), SRMR (.046),

CFI (.97), and NNFI (.96) in the excellent-fit range, these values, along with the AIC (795.34), were substantially better than those for Models 1 and 2 (see Table 3). As shown in Table 4, all items loaded significantly on their respective symptom factors: Emotional Symptoms averaged .59, Conduct Problems averaged .60, Hyperactivity-Inattention Symptoms averaged .68, Peer Problems averaged .50, and Prosocial Behavior averaged .47. Modification indices revealed that model fit would not have been improved substantially if we had allowed items to load on different or multiple factors.

Table 3 also reveals that Model 3 consistently provided the best fit for subsamples defined by grandparent race and grandchild gender and age group, showing configural invariance. Additionally, the more parsimonious models in which factor loadings were constrained to be equal across different levels of these demographic variables generally fit as well as models in which factor loadings were freely estimated for each level, demonstrating metric invariance (see Table 5). Chi-square difference tests provided only slight evidence to the contrary for race, scaled S-B $\chi^2_{\text{diff}}(29, N = 733) = 56.64, p < .01$, and age, scaled S-B $\chi^2_{\text{diff}}(87, N = 733) = 132.64, p < .01$, given the sensitivity of this test to sample size, and they were consistent with other fit indices for gender, scaled S-B $\chi^2_{\text{diff}}(29, N = 733) = 40.53, p > .05$.

In the full sample, 9 of the 10 positively worded items had loadings of .30 or higher on the method factor, with one item loading .19 (averaged .34). At first glance, this suggests that the positively worded items might be reflecting positive construal rather than either strengths or difficulties as intended. However, the symptom factor loadings for 9 of these items were higher than the corresponding method factor loadings, with one item having the same loading on the symptom and method factor (averaged .17 higher on the respective symptom factors than on the method factor). In addition, the item loadings for Prosocial items were on average .10 lower on the Prosocial factor when the method factor was modeled (Model 3) than when it was not modeled (Model 2). The reverse-scored item loadings on the other symptom factors, however, were less affected by the inclusion of the method factor, decreasing only by an average of .05. These descriptive findings suggest that the positively worded Prosocial items might pose more of a method-related threat to the structural validity of the SDQ than do the positively worded items associated with the four problem-oriented symptom dimensions.

We also examined the SDQ's structural validity in terms of the observed pattern of cross-scale correlations and the internal consistency coefficients for each of the subscales for the entire sample. As seen in Table 6, the four symptom subscales were moderately to highly correlated with each other regardless of whether the correlation coefficients were derived error-free via LISREL (range = .64-.78 after modeling the method factor) or by the Pearson product-moment procedure (range = .47-.62). In both cases, the internalizing-externalizing correlations (i.e., Emotional-Conduct; Emotional-Hyperactivity) were of smaller magnitude than the externalizing-externalizing correlations (i.e., Hyperactivity-Conduct). Pearson product-moment correlations of the four symptom scales with the Prosocial subscale were all inverted and ranged from -.30 to -.50. The Total Difficulties score was highly and positively correlated with each symptom subscale (range = .74-.84) but moderately and negatively correlated with the Prosocial subscale (-.50).

Internal consistency as measured by Cronbach's alpha and average interitem correlation (see Table 6) was good for four of the five SDQ subscales (α ranged from .71 to .82; average interitem r ranged from .33 to .48) but was only moderate for the Peer Problems subscale ($\alpha = .62$; average interitem $r = .25$). Cronbach's alpha was .88, and average interitem correlation was .27 for the Total Difficulties score.

Discussion

The overall aim of this study was to examine the structural validity of the parent informant version of the SDQ with a large national sample of custodial grandparents. As found in previous studies involving English-speaking samples (e.g., Bourdon et al., 2005; Goodman, 2001; Hawes & Dadds, 2004), moderate to strong internal reliability was exhibited across all SDQ subscales. Although the observed reliability coefficient for the Peer Problems subscale was less than ideal (.62), this finding is consistent with the low alphas previously found for this subscale. Goodman (2001) reported alpha coefficients as low as .41 (children's self-report) and .57 (parent informant version), whereas Thabet, Stretch, & Vostanis (2000) observed an alpha of .18 among a sample of Arab children. Thus, results from diverse samples indicate potential problems regarding the internal consistency of the SDQ Peer Problems subscale. One explanation for the low internal consistency could be that the scale contains two reverse-scored items (of five items total) that might be contributing measurement error. Removal of these items, however, resulted in a lower estimate of internal consistency, although this could have been due to reliance on only three items to derive the estimate.

The SDQ's structural validity was further evidenced by the pattern of relationships observed among the subscales. Pearson product-moment correlations among all five subscales were statistically significant, but they were low enough in magnitude to suggest distinct constructs (ranging from -.30 to .62). In addition, the strength and direction of these correlations are conceptually meaningful and consistent with current knowledge of comorbidity (Hawes & Dadds, 2004). For example, both error-free correlations obtained from LISREL and Pearson correlations resulted in stronger relationships when the two externalizing subscales (Conduct Problems and Hyperactivity-Inattention) were correlated with one another than when they were correlated with Emotional Symptoms (an internalizing subscale). That the Total Difficulties score was highly correlated with each of the four symptom subscales (range = .74 -.84) supports Goodman's (1997) claim that these subscales can be summed to derive a conceptually valid Total Difficulties score. Inspection of the Pearson product-moment correlations of the 25 SDQ items with each of the five subscales further revealed that each item correlated significantly and most highly with its respective subscale (data available on request). Not only are these findings consistent with those obtained in prior English-speaking samples (Goodman, 2001; Hawes & Dadds, 2004), but ours is the first study to confirm these conceptually meaningful relationships among SDQ subscales using SEM.

Our CFA model fit results are supportive of the proposed SDQ factor structure. The fact that Models 1 and 2, which are slightly different variants of Goodman's (1997) predicted five-factor structure, both demonstrated adequate model fit suggests that his original conceptualization of the SDQ was upheld in the present sample of custodial grandparents. These findings are consistent with prior studies of English-speaking parent informants in which Goodman's (1997) predicted five-factor structure was confirmed via PCA with minimal cross-loadings observed between the subscales (Goodman, 2001; Hawes & Dadds, 2004).

Our use of CFA in the present study, however, permitted us to extend these prior findings in three key ways. First, the finding that Model 2 fit the data significantly better than Model 1 suggests that each of the four symptom subscales is better viewed as a distinct measure in its own right rather than as a mere indicator of overall psychopathology, as might otherwise be inferred from the Total Difficulties score. Nevertheless, the extent of covariance observed among these four subscales was of sufficient magnitude to warrant their summation to derive a Total Difficulties score. Rather than implying that the summed Total Difficulties score is meaningless and inappropriate, the superiority of the correlated four-factor model relative to the hierarchical four-factor model highlights the fact that it is also important to recognize independently the specific symptom dimensions that the overall behavioral difficulties

construct comprises. Second, the finding that applying equality constraints on the factor loadings across respondent race, target gender, and target age group did not substantially diminish model fit demonstrates that the factor structure of the SDQ holds across these demographic variables. Third, the findings associated with Model 3 provide evidence that a positive construal factor does taint the SDQ, as previously speculated (Dickey & Blumberg, 2004; Goodman, 1997). However, on a positive note, our results also suggest that the impact of any response bias arising from this factor is negligible with respect to the four SDQ symptom sub-scales and is limited chiefly to the Prosocial subscale.

Despite these findings, there are at least three reasons why it is premature to abandon the Prosocial items in order to shorten the SDQ. First, the present findings could be unique to custodial grandparents as informants and thus may not generalize to other informant populations (e.g., parents, teachers, youths). Second, Goodman (1997) maintained that the equivalence that has been consistently found between Rutter and SDQ scores indicates that the Prosocial subscale does not exert an adverse effect on the SDQ symptom subscales. Third, Dickey and Blumberg (2004) cautioned that removing the Prosocial items could make the instrument less acceptable to respondents and result in the underreporting of behavioral difficulties. Moreover, because we did not administer alternative versions of the SDQ with and without the Prosocial items, it is impossible to determine whether removal of these items might have any adverse effect on the remaining problem-focused items.

It is important to note that, with the exception of the evidence supporting the positive construal method factor, our overall CFA findings are discrepant from the only other published study of the SDQ's factor structure with a U.S. sample. Dickey and Blumberg (2004) found that parental informants reported on three separate but correlated underlying dimensions: externalizing problems (Hyperactivity-Inattention and Conduct Problems), internalizing problems (Emotional Problems and Peer Problems), and a positive construal method factor. The CFA support for their model, however, was obtained from a modified SDQ in which one item was deleted, not on theoretical grounds, but on the basis of sample-specific PCA and EFA results. Given our interest in the complete SDQ and in testing Goodman's theoretical model and given the inability to directly compare models based on different sets of variables, we chose not to test this three-factor model.

Several other possible explanations exist for the discrepant findings between the two studies. First, only a small fraction of Dickey and Blumberg's (2004) respondents were grandparents, and our findings may be unique to custodial grandparents as informants. Second, Dickey and Blumberg used the new Americanized SDQ, whereas we used the older British English version. Third, data in Dickey and Blumberg were obtained from NHIS face-to-face interviews, whereas our data were obtained from telephone interviews. Fourth, the NHIS data were based on a representative sample of all races, whereas our data were from Blacks and Whites only. Fifth, different analytic approaches were used in the two studies. Whereas the CFA conducted by Dickey and Blumberg was based on the findings of preceding PCA and EFA and excluded one of the Conduct Problem items, we opted to use only CFA as an explicit test of several plausible structural models.

In summary, our findings with respect to (a) internal consistency estimates for individual subscales, (b) the observed pattern of cross-scale correlations, (c) the CFA results across three separate models, and (d) the measurement equivalence across demographic variables are highly supportive of the SDQ's factor structure in the present sample of U.S. custodial grandparents. These findings are important because this is the first study to test the psychometric properties of the SDQ with informants other than parents, teachers, or children. Given that increasing numbers of grandparents are assuming full-time parental responsibility for their grandchildren, it is reassuring to know that measures of the grandchild's psychological adjustment may be

meaningfully administered to this expanding population of surrogate parents. The significance of this is clearly signaled by the fact that over 5% of the respondents in the 2001 NHIS were relatives other than parents.

Given that our data were obtained by means of a telephone interview, the present findings also provide initial evidence that the SDQ may be effectively administered in formats other than traditional face-to-face interviews or pencil-and-paper questionnaires. Our phone administration of the SDQ was accepted well by custodial grandparents, interviewers did not report any administrative problems, and all of the resulting SDQ data were useable. Not only are these findings consistent with past research on the formats of administering clinical measures (Baer et al., 1997), they also support the continued use of the SDQ in studies that rely on telephone interviews for gathering data.

Study Limitations

Several limitations of the present study must be acknowledged. First, because our sample was limited to Black and White respondents, future research with additional racial and ethnic groups (e.g., Latinos, Asians, and Native Americans) is necessary. Second, several important clinical and psychometric properties of the SDQ (e.g., the ability to identify “caseness,” diagnostic concordance, clinical cutoff scores, test-retest reliability, sensitivity as an outcome measure, and associations with external variables) were not examined. Such goals were beyond the scope of the present study, in which relevant clinical data and evaluations from mental health professionals were not obtained. Future research on these issues is needed before definite conclusions regarding the usefulness of the SDQ with custodial grandparents can be reached. Third, despite a large national sample, the extent to which the present sample was representative of the total U.S. population of custodial grandparents and their grandchildren is unknown. Finally, the present study focused solely on the parent informant version of the SDQ, with custodial grandmothers as respondents.

Future Research Directions and Conclusions

An important goal of future research is to examine the clinical usefulness and psychometric properties of the SDQ when it is administered to other potential informants such as grandfathers, teachers, and grandchildren themselves in regard to the psychological adjustment of custodial grandchildren. Despite its limitations, the present study provides an initial confirmation of the SDQ’s purported factor structure (Goodman, 1997), at least in terms of problem-oriented factors, with both Black and White custodial grandmothers responding as nontraditional informants.

Acknowledgements

Work on this study was funded by National Institute of Mental Health Grant R01MH66851-02 awarded to Gregory C. Smith.

References

- Achenbach, T.M. Manual for the Child Behavior Checklist/4-18 and 1991 Profile. University of Vermont Department of Psychiatry; Burlington, VT: 1991.
- Achenbach, T.M.; Howell, C.T.; Quay, H.C.; Conners, C.K. National survey of competencies and problems among 4- to 16-year-olds: Parents’ reports for normative and clinical samples. *Monographs of the Society for Research in Child Development* 1991;56(3) Serial No. 225 1131 [PubMed: 1770964]
- Akaike, H. Factor analysis and AIC. *Psychometrika* 1987;52:317–332.
- Baer, L.; Cukor, P.; Coyle, J.T. Telepsychiatry: Application of telemedicine to psychiatry. In: Bashid, R.; Sanders, J.; Shannon, G., editors. *Telemedicine: Theory and practice*. Charles C. Thomas; Springfield, IL: 1997.

- Becker A, Woerner W, Hasselhorn M, Banaschewski T, Rothenberger A. Validation of the parent and teacher SDQ in a clinical sample. *European Child and Adolescent Psychiatry* 2004;13(Suppl 2):11–16.
- Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin* 1990;107:238–246. [PubMed: 2320703]
- Bentler PM, Bonett D. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin* 1980;88:588–606.
- Bentler PM, Dudgeon P. Covariance structure analysis: Statistical practice, theory, and directions. *Annual Review of Psychology* 1996;47:563–592.
- Bollen, KA. *Structural equations with latent variables*. Wiley; New York: 1989.
- Bourdon KH, Goodman R, Rae D, Simpson G, Koretz DS. The Strengths and Difficulties Questionnaire: U.S. normative data and psychometric properties. *Journal of the American Academy of Child and Adolescent Psychiatry* 2005;44:557–564. [PubMed: 15908838]
- Bratton S, Ray D, Moffit K. Filial/family play therapy: An intervention for custodial grandparents and their grandchildren. *Educational Gerontology* 1998;24:391–406.
- Browne MW, Cudeck R. Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research* 1989;24:445–455.
- Brown-Standridge MD, Floyd CW. Healing bittersweet legacies: Revisiting contextual family therapy for grandparents raising grandchildren in crisis. *Journal of Marriage and Family Counseling* 2000;26:180–197.
- Browne, MW.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA.; Long, JS., editors. *Testing structural equation models*. Sage; Beverly Hills, CA: 1993. p. 136-162.
- Dickey WC, Blumberg SJ. Revisiting the factor structure of the Strengths and Difficulties Questionnaire: United States, 2001. *Journal of the American Academy of Child and Adolescent Psychiatry* 2004;43:1159–1167. [PubMed: 15322420]
- Ghuman HS, Weist MD, Shafer ME. Demographic and clinical characteristics of emotionally disturbed children being raised by grandparents. *Psychiatric Services* 1999;50:1496–1498. [PubMed: 10543863]
- Goodman R. The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry* 1997;38:581–586. [PubMed: 9255702]
- Goodman R. The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. *Journal of Child Psychology and Psychiatry* 1999;40:791–801. [PubMed: 10433412]
- Goodman R. Psychometric properties of the Strengths and Difficulties Questionnaire (SDQ). *Journal of the American Academy of Child and Adolescent Psychiatry* 2001;40:1337–1345. [PubMed: 11699809]
- Goodman R, Ford T, Simmons H, Gatward R, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *British Journal of Psychiatry* 2000;177:534–539. [PubMed: 11102329]
- Goodman R, Meltzer H, Bailey V. The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. *International Review of Psychiatry* 2003;15:173–177. [PubMed: 12745329]
- Gorsuch, RL. *Factor analysis*. 2nd ed.. Erlbaum; Hillsdale, NJ: 1983.
- Hawes DJ, Dadds MR. Australian data and psychometric properties of the Strengths and Difficulties Questionnaire. *Australian and New Zealand Journal of Psychiatry* 2004;38:644–651. [PubMed: 15298588]
- Hayslip B, Shore RJ, Henderson CE, Lambert PL. Custodial grandparenting and the impact of grandchildren with problems on role satisfaction and role meaning. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences* 1998;53:S164–S173.
- Hirshorn, B. Grandparents as caregivers. In: Szinovacz, M., editor. *Handbook on grandparenthood*. Greenwood Press; Westport, CT: 1998. p. 200-214.
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling* 1999;6:1–55.

- Jöreskog, KG.; Sörbom, D. New features in LISREL 8. Scientific Software; Chicago: 1993.
- Jöreskog, KG.; Sörbom, D.; du Toit, S.; du Toit, M. LISREL 8: New statistical features. Scientific Software International; Lincolnwood, IL: 2000.
- Kish, L. Survey sampling. Wiley; New York: 1965.
- Koskelainen M, Sourander A, Kaljonen A. The Strengths and Difficulties Questionnaire among Finnish school-aged children and adolescents. *European Child and Adolescent Psychiatry* 2000;9:277–284. [PubMed: 11202103]
- Koskelainen M, Sourander A, Vauras M. Self-reported strengths and difficulties in a community sample of Finnish adolescents. *European Child and Adolescent Psychiatry* 2001;10:180–185. [PubMed: 11596818]
- Malmberg M, Rydell AM, Smedje H. Validity of the Swedish version of the Strengths and Difficulties Questionnaire (SDQ-Swe). *Nordic Journal of Psychiatry* 2003;57:357–363. [PubMed: 14522609]
- Marzocchi G, Capron C, Di Pietro M, Tauleria E, Duyme M, Frigerio A, et al. The use of the Strengths and Difficulties Questionnaire (SDQ) in Southern European countries. *European Child and Adolescent Psychiatry* 2004;13(Suppl 2):40–46.
- Mathai J, Anderson P, Bourne A. Use of the Strengths and Difficulties Questionnaire as an outcome measure in a child and adolescent mental health service. *Australasian Psychiatry* 2003;11:334–337.
- Muris P, Meesters C, Eijkelenboom A, Vincken M. The self-report version of the Strengths and Difficulties Questionnaire: Its psychometric properties in 8- to 13-year-old non-clinical children. *British Journal of Clinical Psychology* 2004;43:437–448. [PubMed: 15530213]
- Muris P, Meesters C, van den Berg F. 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* 2003;12:1–8. [PubMed: 12601558]
- National Center for Health Statistics. National Health Interview Survey (NHIS). U.S. Centers for Disease Control and Prevention; Hyattsville, MD: 2003.
- Obel C, Heiervang E, Rodriguez A, Heyerdahl S, Smedje H, Sourander A, et al. The Strengths and Difficulties Questionnaire in the Nordic countries. *European Child and Adolescent Psychiatry* 2004;13(Suppl 2):32–39.
- Pebley AR, Rudkin LL. Grandparents caring for grandchildren: What do we know? *Journal of Family Issues* 1999;20:218–242.
- Rutter MA. Children's behavior questionnaire for completion by teachers: Preliminary findings. *Journal of Child Psychology and Psychiatry* 1967;8:1–11. [PubMed: 6033260]
- Satorra, A.; Bentler, PM. Scaling corrections for chi-square statistics in covariance structure analysis; ASA Proceedings of the Business and Economic Statistics Section; Alexandria, VA: American Statistical Association. 1988; p. 308-313.
- Satorra A, Bentler PM. A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika* 2001;66:507–514.
- Schumacker RE, Beyerlein ST. Confirmatory factor analysis with different correlation types and estimation methods. *Structural Equation Modeling* 2000;7:629–636.
- Smedje H, Broman E, Hetta J, von Knorring AL. Psychometric properties of a Swedish version of the Strengths and Difficulties Questionnaire. *European Child and Adolescent Psychiatry* 1999;8:63–70. [PubMed: 10435454]
- Smith GC, Savage-Stevens S, Fabian E. How caregiving grandparents view support groups for grandchildren in their care. *Family Relations* 2002;51:274–281.
- Steiger JH. Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research* 1990;25:173–180.
- Thabet AA, Stretch D, Vostanis P. Child mental health problems in Arab children: Application of the Strengths and Difficulties Questionnaire. *International Journal of Social Psychiatry* 2000;46:266–280. [PubMed: 11201348]
- U.S. Bureau of the Census. Profile of general demographic characteristics: 2000. U.S. Author; Washington, DC: 2004.

- van Widenfelt BM, Goedhart AW, Treffers PDA, Goodman R. Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *European Child and Adolescent Psychiatry* 2003;12:281–289. [PubMed: 14689260]
- Woerner W, Becker A, Rothenberger A. Normative data and scale properties of the German parent SDQ. *European Child and Adolescent Psychiatry* 2004;13(Suppl 2):3–10. [PubMed: 15692877]
- Woerner W, Fleitlich-Bilyk B, Martinussen R, Fletcher J, Cucchiaro G, Dalgarrondo P, et al. The Strengths and Difficulties Questionnaire overseas: Evaluations and applications of the SDQ beyond Europe. *European Child and Adolescent Psychiatry* 2004;13(Suppl 2):40–46.

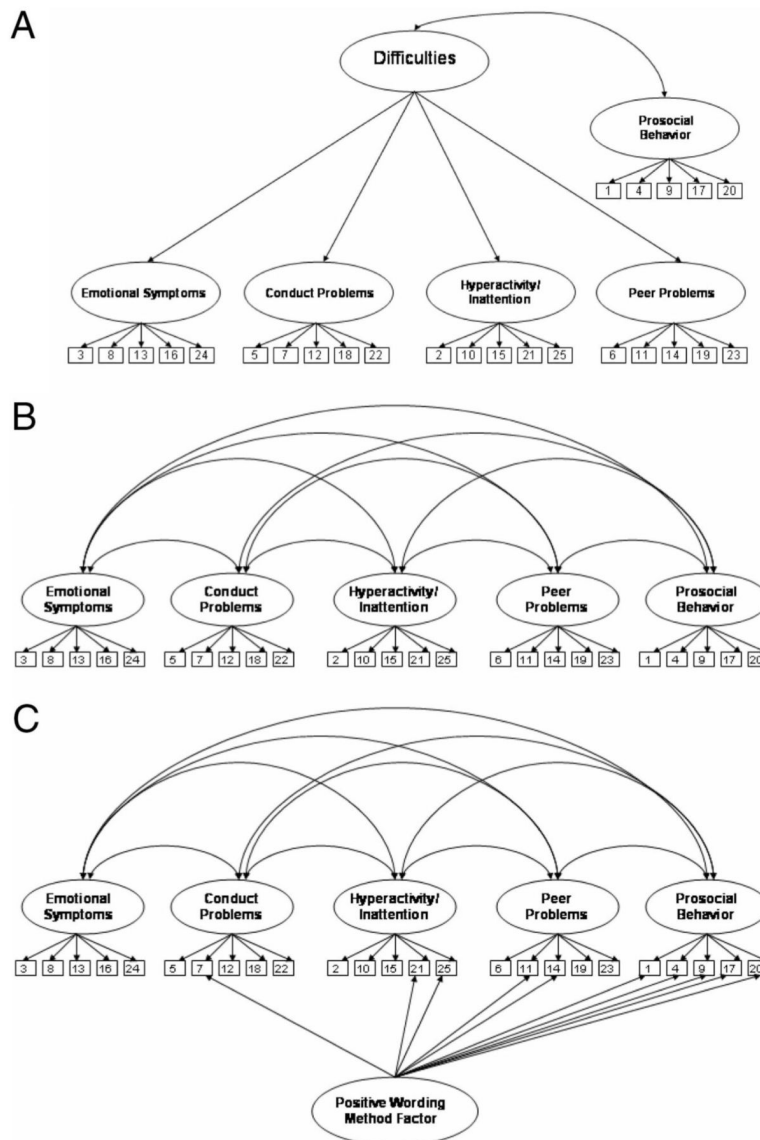


Figure 1. Three hypothesized structural models of the Strength and Difficulties Questionnaire were evaluated by means of confirmatory factor analysis. A: Five-factor higher order model (Model 1). B: Five-factor lower order model (Model 2). C: Six-factor model, including a positive wording method factor (Model 3). Disturbances are omitted to simplify the graphical depictions.

Table 1
Descriptive Demographic Data for Grandmothers and Target Grandchildren (N = 733)

Variable	n	%	Variable	n	%
Race			Race		
White	367	50.1	Black	366	49.9
Marital status			Marital status		
Married	352	48.0	Divorced	159	21.7
Widowed	102	13.9	Single, never married	66	9.0
Separated	49	6.7	Living with partner (not married)	5	0.7
Education			Education		
Less than 5 years	5	0.7	Some college	260	35.5
5-8 years	19	2.6	College graduate	95	13.0
Some high school	99	13.5	Graduate school and/or professional training	50	6.8
High school graduate	205	28.0			
Residential locale			Residential locale		
Urban	350	47.8	Rural	238	32.5
Suburban	141	19.2	Other	4	0.5
Work status			Work status		
Not working	202	27.6	Retired	144	19.6
Working full- or part-time	329	44.9	Full-time homemaker	41	5.6
Seeking employment	17	2.3			
Income			Income		
Under \$10,000	123	16.8	\$36,000-\$50,000	99	13.5
\$10,000-\$15,000	109	14.9	\$51,000-\$75,000	59	8.0
\$16,000-\$20,000	85	11.6	\$76,000-\$100,000	19	2.6
\$21,000-\$25,000	90	12.3	\$101,000-\$125,000	4	0.5
\$26,000-\$35,000	112	15.3	More than \$125,000	5	0.7
			Missing data	28	3.8
Target grandchild age ^a (years)			Target grandchild age ^a (years)		
4-7	235	32.1	11-14	206	28.1
8-10	194	26.5	15-17	98	13.4
Formal legal arrangement			Formal legal arrangement		
Formal custody	271	37.0	Foster parent status	16	2.2
Adoption	90	12.3	Currently seeking legal custody	69	9.4
Guardianship	215	29.3	None	72	9.8
Relationship of target grandchild's parent to grandmother			Relationship of target grandchild's parent to grandmother		
Son	227	31.0	Daughter	482	65.8
Stepson	7	1.0	Stepdaughter	16	2.2
Missing data	1	0.1			
Reasons for care ^b			Reasons for care ^b		
Child abandonment	205	28.0	Parental substance abuse	406	55.4
Physical or emotional abuse	201	27.4	Parents' incarceration	312	42.6
Removed from parental custody	225	30.7	Parents' HIV/AIDS	11	1.5
Parents' mental health problems	218	29.7	Other (not AIDS) health problems	114	15.6
Parents' teen pregnancy	133	18.1	One of the parents deceased	47	6.4

^a Age groupings shown for target grandchildren are the same as those presented for the Strengths and Difficulties Questionnaire normative sample (see www.sdqinfo.com).

^b Respondents were asked to report any and all reasons for assuming care of the target grandchild that applied to their situation.

Table 2
SDQ Item Means, Standard Deviations, Skewness, and Kurtosis for the Full Sample (N = 733)

SDQ abbreviated item content	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
1. Considerate of others	1.39	0.67	-0.63	-0.67
2. Restless, overactive	0.87	0.83	0.25	-1.51
3. Somatic complaints	0.40	0.65	1.38	0.62
4. Shares with others	1.40	0.71	-0.76	-0.69
5. Temper tantrums	0.78	0.81	0.43	-1.36
6. Rather solitary	0.60	0.76	0.82	-0.82
7. <i>Generally obedient</i>	0.69	0.66	0.43	-0.75
8. Many worries	0.58	0.73	0.85	-0.65
9. Helpful	1.60	0.63	-1.33	0.60
10. Constantly fidgeting	0.64	0.80	0.74	-1.04
11. <i>Has a good friend</i>	0.26	0.57	2.07	3.11
12. Often fights, bullies	0.41	0.66	1.34	0.49
13. Often unhappy	0.44	0.64	1.18	0.23
14. <i>Liked by others</i>	0.35	0.58	1.45	1.09
15. Easily distracted	0.92	0.82	0.16	-1.49
16. Nervous, clingy	0.66	0.79	0.68	-1.08
17. Kind to younger kids	1.73	0.53	-1.81	2.36
18. Often lies, cheats	0.52	0.71	0.98	-0.38
19. Bullied by others	0.46	0.66	1.14	0.07
20. Volunteers to help	1.50	0.64	-0.90	-0.26
21. <i>Thinks before acting</i>	1.11	0.73	-0.17	-1.10
22. Steals	0.19	0.49	2.56	5.71
23. Gets along better with adults than children	0.75	0.77	0.46	-1.16
24. Many fears	0.53	0.73	0.98	-0.45
25. <i>Good attention span</i>	0.90	0.77	0.17	-1.31

Note. SDQ = Strengths and Difficulties Questionnaire. Items in italics are reverse-scored for appropriate keying to the relevant factor. Tests of skewness and kurtosis simultaneously were significant for each item at $p < .001$.

Table 3

Fit Statistics for Robust Maximum-Likelihood Confirmatory Factor Analyses

Model	df	χ^2	S-B χ^2	RMSEA (90% CI)	SRMR	CFI	NNFI	AIC
Full sample (N = 733)								
1	270	1048.96	967.84	.059 (.055, .063)	.059	.95	.94	1077.84
2	265	991.29	909.71	.058 (.054, .062)	.055	.95	.94	1029.71
3	255	761.20	655.34	.046 (.042, .051)	.046	.97	.96	795.34
Grandparent race								
Black (n = 366)								
1	270	692.55	627.97	.060 (.054, .066)	.066	.94	.93	737.97
2	265	668.51	605.73	.059 (.053, .066)	.064	.94	.93	725.73
3	255	517.13	429.94	.043 (.036, .050)	.051	.97	.96	569.94
White (n = 367)								
1	270	693.79	626.35	.060 (.054, .066)	.066	.95	.95	736.35
2	265	652.69	583.53	.057 (.051, .064)	.062	.96	.95	703.53
3	255	561.81	489.45	.050 (.043, .057)	.056	.97	.96	629.45
Grandchild gender								
Female (n = 391)								
1	270	717.91	630.00	.058 (.053, .064)	.065	.94	.94	740.00
2	265	685.96	597.44	.057 (.051, .063)	.062	.95	.94	717.44
3	255	540.37	434.87	.043 (.036, .049)	.054	.97	.97	574.87
Male (n = 342)								
1	270	646.93	580.86	.058 (.052, .065)	.064	.95	.95	690.86
2	265	618.20	557.83	.057 (.050, .064)	.060	.96	.95	677.83
3	255	512.09	443.22	.047 (.039, .054)	.050	.97	.97	583.22
Grandchild age group								
4-7 years (n = 235)								
1	270	476.21	426.75	.050 (.041, .059)	.069	.96	.95	536.75
2	265	457.06	408.33	.048 (.039, .057)	.066	.96	.96	528.33
3	255	381.09	334.12	.036 (.025, .047)	.058	.98	.98	474.12
8-10 years (n = 194)								
1	270	563.47	544.49	.073 (.064, .081)	.078	.91	.90	654.49
2	265	551.93	527.98	.072 (.063, .081)	.077	.92	.90	647.98
3	255	442.81	386.23	.052 (.041, .062)	.068	.96	.95	526.23
11-14 years (n = 206)								
1	270	600.18	490.73	.063 (.054, .072)	.075	.95	.94	600.73
2	265	583.31	486.19	.064 (.055, .073)	.071	.95	.94	606.19
3	255	544.98	459.27	.063 (.053, .072)	.070	.95	.95	599.27
15-17 years (n = 98)								
1	270	468.20	361.12	.059 (.042, .074)	.094	.95	.94	471.12
2	265	440.67	333.74	.052 (.032, .068)	.089	.96	.96	453.74
3	255	398.12	296.73	.041 (.011, .060)	.081	.98	.97	436.73

Note. df = degrees of freedom; S-B = Satorra-Bentler; RMSEA = root-mean-square error of approximation; CI = confidence interval; SRMR = standardized root-mean-square residual; CFI = comparative fit index; NNFI = nonnormed fit index; AIC = Akaike information criterion.

Table 4
 Completely Standardized Factor Loadings for First-Order SDQ Model Before (Model 2) and After (Model 3) Accounting for the Effects of the Method Factor (N = 733)

SDQ item	Emotional Problems	SDQ item	Conduct Problems	SDQ item	Hyperactivity-Inattention	SDQ item	Peer Problems	SDQ item	Prosocial Behavior
3	.48/.48	5	.71/.72	2	.76/.77	6	.48/.51	1	.68/.57
8	.69/.68	7	.62/.57	10	.74/.76	11	.46/.40	4	.61/.52
13	.67/.67	12	.66/.67	15	.73/.73	14	.65/.60	9	.42/.30
16	.62/.62	18	.62/.63	21	.54/.48	19	.51/.59	17	.54/.52
24	.51/.51	22	.39/.39	25	.69/.65	23	.39/.41	20	.61/.46

Note. SDQ = Strengths and Difficulties Questionnaire. For ease of interpretation, items were ordered by factor. Loadings before and after the slash are based on Models 2 and 3, respectively.

Model 3 Fit Statistics for Configural Invariance and Metric Invariance Models for Grandparent Race and Grandchild Gender and Age Group

Table 5

Model	df	χ^2	S-B χ^2	RMSEA (90% CI)	CFI	NNFI	AIC
Race							
CINV	510	1,078.95	918.91	.047 (.042, .052)	.97	.97	1,198.91
MINV	539	1,138.04	975.48	.047 (.042, .052)	.97	.97	1,197.48
Gender							
CINV	510	1,052.45	877.75	.044 (.039, .049)	.98	.97	1,157.75
MINV	539	1,096.38	918.18	.044 (.039, .049)	.97	.97	1,140.18
Age							
CINV	1020	1,767.00	1474.55	.049 (.044, .055)	.97	.97	2,034.55
MINV	1107	1,932.04	1607.58	.050 (.044, .055)	.97	.96	1,933.58

Note. df = degrees of freedom; S-B = Satorra-Bentler; RMSEA = root-mean-square error of approximation; CI = confidence interval; CFI = comparative fit index; NNFI = nonnormed fit index; AIC = Akaike information criterion; CINV = configural invariance, unconstrained factor loadings; MINV = metric invariance, constrained factor loadings.

SDQ Scale Descriptive Statistics, Pearson Product-Moment Correlations, and Factor Correlations for the Full Sample (N = 733)

Table 6

SDQ subscale	1	2	3	4	5	M	SD	α	Average interitem correlation
1. Emotional Symptoms	—	.67	.64	.78	-.55	2.61	2.45	.73	.35
2. Conduct Problems	.48	—	.76	.68	-.76	2.60	2.37	.74	.36
3. Hyperactivity-Inattention	.49	.62	—	.65	-.48	4.43	3.02	.82	.48
4. Peer Problems	.52	.45	.47	—	-.64	2.42	2.12	.62	.25
5. Prosocial Behavior	-.30	-.50	-.38	-.40	—	7.61	2.18	.71	.33
6. Total Difficulties	.78	.80	.84	.74	-.50	12.05	7.92	.88	.27

Note. Pearson product-moment correlations are below the diagonal; substantive factor correlations from Model 3 are above the diagonal. Each subscale has five items and an observed range of 0 to 10; the SDQ Total Difficulties scale excludes the five Prosocial Behavior items and has an observed range of 0 to 35.