AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Assessment

ISSN: 1040-3590

2021, Vol. 33, No. 1, 45–59
http://dx.doi.org/10.1037/pas0000961

The Strengths and Difficulties Questionnaire: An Examination of Factorial, Convergent, and Discriminant Validity Using Multitrait-Multirater Data

Tiago Ferreira
University of Porto

Christian Geiser
Utah State University

Joana Cadima, Marisa Matias, Teresa Leal, and Paula Mena Matos
University of Porto

This study examined the factorial, convergent, and discriminant validity of scores on the Strengths and Difficulties Questionnaire (SDQ), a tool for screening children's psychological adjustment. Data were collected from a community sample of 346 children (46% girls, *M* age = 54.07 months), using teacher, mother, and father reports. Results from confirmatory factor analyses indicated that the SDQ's hypothesized 5-factor structure fit the data well and partial strict measurement invariance was established across raters. Using teachers' reports as reference method, a correlated trait–correlated method minus 1 model (Eid et al., 2008) was fitted to investigate convergent and discriminant validity. The convergent validity of parents' ratings relative to teachers' ratings was modest. Mothers and fathers had a unique perspective on children's behavior above and beyond their partial overlap with teacher reports. Results indicated good discriminant validity between most of the traits measured by the SDQ.

Public Significance Statement
This study investigates the ability of the Strengths and Difficulties Questionnaire to discriminate between different traits and the degree of convergence between teacher, mother, and father ratings. Among other implications, our findings confirm the importance of using multiple informants to more comprehensively assess children's behavior, stressing the need for considering reports based on different observation contexts, namely home and school.

Keywords: Strengths and Difficulties Questionnaire, multitrait-multimethod, convergent validity, discriminant validity, correlated trait-correlated method minus one model

Supplemental materials: http://dx.doi.org/10.1037/pas0000961.supp

The Strengths and Difficulties Questionnaire (SDQ) is a widely used measure of children's psychological adjustment, offering the possibility of screening behavior through multiple perspectives (Goodman, 1997). This short questionnaire was designed to assess children between 4 and 16 years of age using comparable versions for collecting parent reports, teacher reports, and child self-ratings. Although the possibility of assessing child behavior through multiple perspectives is one of the SDQ's main assets, there is limited evidence regarding the equivalence of the psychometric properties of this measure across rater types. The current study focused on preschool-aged children, examining the invariance of the measurement properties of this measure across ratings from teachers, mothers, and fathers. We also examined the SDQ's ability to discriminate between different traits and the degree of convergence across raters when reporting the same trait using a multitrait-multimethod (MTMM) approach. This study addresses the convergence between ratings from multiple informants collected through the same method (i.e., questionnaire). We adopt the term "multitrait-multimethod" throughout the article as this term commonly refers to the general analytical framework used to investigate the discriminant and convergent validity of scores obtained from multiple methods or raters.

The Strengths and Difficulties Questionnaire (SDQ)

The SDQ was designed by R. Goodman (1997) to measure children's psychological adjustment. This brief questionnaire includes a small number of items that can be completed by raters

This article was published Online First October 29, 2020.

Tiago Ferreira, Faculty of Psychology and Education Sciences, Center for Psychology, University of Porto; Christian Geiser, Department of Psychology, Utah State University; Joana Cadima, Marisa Matias, Teresa Leal, and Paula Mena Matos, Faculty of Psychology and Education Sciences, Center for Psychology, University of Porto.

This research was supported by FEDER, COMPETE program, and by the Portuguese Foundation for Science and Technology, PTDC/MHC-CED/5218/2012 and PD/BD/114269/2016 grants.

Correspondence concerning this article should be addressed to Tiago Ferreira, Faculdade de Psicologia e de Ciências da Educação, Universidade do Porto, Rua Alfredo Allen 4200-135 Porto, Portugal. E-mail: tiagodsferreira@gmail.com

from different educational, social, and cultural backgrounds (Marzocchi et al., 2004; Stolk, Kaplan, & Szwarc, 2017; Williamson et al., 2014). The 25 items are equally divided among five dimensions of children's social, emotional, and behavioral functioning, namely emotional symptoms (ES), peer problems (PP), conduct problems (CP), hyperactivity (Hy), and prosocial behaviors (PB). These dimensions can be reported by parents and teachers of children aged 4 to 16 years using the same version of the questionnaire and self-reported by older children aged 11 to 16 years using a slightly adapted version.

The SDQ covers a wide age range, is translated into over 40 languages, has normative data from several countries and is freely available for download (see https://sdqinfo.org). These characteristics make this instrument one of the most popular tools for screening children's psychological adjustment. Besides its utility for clinical practice, the SDQ is well-suited for conducting research in community and clinical samples (Smits, Theunissen, Reijneveld, Nauta, & Timmerman, 2018) and is frequently used in large epidemiological surveys (Kremer et al., 2015) as well as cross-country comparative research (Vries, Davids, Mathews, & Aarø, 2018).

The SDQ's Portuguese translation used in this study has demonstrated adequate psychometric properties across different studies conducted in Portugal (Marzocchi et al., 2004). Overall, results from studies using the Portuguese version support the SDQ's construct validity, showing a factorial structure and pattern of loadings that match those found in the United Kingdom and other countries, namely Southern European countries (Marzocchi et al., 2004). Nevertheless, further work is needed to fully understand whether the different translations of the SDQ available are culturally equivalent (Kersten et al., 2016).

Factor Structure and Measurement Invariance

Since its release, a large number of studies around the world have scrutinized the SDQ's measurement properties (see, Kersten et al., 2016 as well as Stone, Otten, Engels, Vermulst, & Janssens, 2010, for reviews). Specifically, the SDQ's factorial validity has been investigated through confirmatory factor analysis (CFA), a procedure used to examine the degree to which the scores of an instrument reflect the hypothesized dimensions (factors) of the construct under evaluation (Jöreskog, 1969). For the parent, teacher, and self-rated versions, a number of empirical studies have supported the five-factor model originally proposed by Goodman (1997; Klein, Otto, Fuchs, Zenger, & von Klitzing, 2013; Niclasen, Skovgaard, Andersen, Sømhovd, & Obel, 2013; van Roy, Veenstra, & Clench-Aas, 2008). Nevertheless alternative factor models can be found in the literature (Caci, Morin, & Tran, 2015; Goodman, Lamping, & Ploubidis, 2010; Kóbor, Takács, & Urbán, 2013; Palmieri & Smith, 2007). Some studies suggested the usefulness of including second-order internalizing and externalizing factors for evaluating low-risk samples (Goodman et al., 2010), though they frequently retain the model with five first-order factors as the best fitting model (Klein et al., 2013; Van Leeuwen, Meerschaert, Bosmans, de Medts, & Braet, 2006). Also, distinct bifactor models, including one or more global factors in addition to the five original factors, have been confirmed as the best-fitting model (Caci et al., 2015; Kóbor et al., 2013; Palmieri & Smith, 2007). However, results from some of these studies suggested that the differences between the original and the bifactor, less parsimonious models may not be significant across different samples (Caci et al., 2015; Palmieri & Smith, 2007). Specifically, for young children, most of the previous studies support the five-factor model, for both the parent and teacher versions (Klein et al., 2013; Mieloo et al., 2012; Niclasen et al., 2013; Sanne, Torsheim, Heiervang, & Stormark, 2009; Van Leeuwen et al., 2006).

Despite the wide empirical support for the SDQ's five-factor model, there is a relatively scarce number of studies examining whether the SDQ's psychometric properties are comparable across different raters/methods (Niclasen et al., 2013; Rogge, Koglin, & Petermann, 2018; Sanne et al., 2009). The examination of measurement invariance (MI; aka measurement equivalence) allows determining the degree to which a set of items or scales reflects constructs in the same or similar way across raters and is of special interest in multitrait-multimethod (MTMM) studies that use different rater types (Geiser, Burns, & Servera, 2014). MI is a prerequisite for meaningful comparisons of structural parameters (e.g., latent factor means and variances) across reporter types.

Formally, MI means that (a) the same basic factor structure (number of factors and patterns of loadings) is found across raters; and/or that (b) factor loadings (multiplicative constants), intercepts (additive constants), and/or measurement error variances take on the same values for a given measure across reporters (Widaman & Reise, 1997). An equivalent factor structure without formal equality of loadings, intercepts, or error variances is known as the condition of *configural invariance*. If, in addition to configural invariance, equal factor loadings are found across reporters, this is referred to as *weak* or *metric invariance*. Weak invariance plus equal intercepts constitutes the condition of *strong* or *scalar invariance*.

When strong invariance holds, this means that the latent factors are measured on comparable scales across reporters, that is, with the same origin and units of measurement. As a consequence, the factor means and variances can be meaningfully compared across raters. This is of interest to studies wanting to compare mean levels and/or true variability across different reporters. Strict measurement invariance requires equal measurement error variances in addition to the conditions implied by strong invariance. Strict invariance is not required for a meaningful comparison of latent factor means and variances across reporters. Nonetheless, strict invariance is an interesting condition because it implies equal measurement precision across reporters.

To our knowledge, the only study supporting the SDQ's strict invariance across raters was conducted by Rogge, Koglin, and Petermann (2018). Using a sample of 3- to 6-year-old children, their results suggested that strict MI is tenable across teacher and parent (mother or father) ratings. Sanne et al. (2009) established partial metric invariance using a large sample of school-age children, whereas results from Niclasen, Skovgaard, Andersen, Sømhovd, and Obel (2013) using four population based cohorts of preschool and school-age children did not support metric invariance between parents and teacher ratings. The inconsistent findings regarding the level of MI of the SDQ across parents and teacher reports might be related to sample differences related to children's age and sex. Although Niclasen et al. (2013) did not find differences between boys and girls, their results suggest that the

SDQ factor structure provides a better fit to data from older children (10 to 12 years of age) than to data from younger children (5 to 7 years of age).

Furthermore, some studies used mothers to represent parent ratings (Niclasen et al., 2013; Sanne et al., 2009) whereas others used either mother or father ratings to represent parents reports (Rogge et al., 2018). Thus, different conclusions regarding the degree of MI between parents' and teacher ratings might have stemmed from different ways of defining and measuring parent ratings. Despite their unique perspectives on children's behavior (Davé, Nazareth, Senior, & Sherr, 2008), to our knowledge, no previous study has examined the invariance of the SDQ scores across mothers and fathers reports. The current study aimed to clarify the extent to which the psychometric properties of the SDQ items are the same across mother, father, and teacher reports.

Convergent and Discriminant Validity

Campbell and Fiske (1959) proposed the MTMM analytical framework to examine the convergent and discriminant validity of a set of measures based on a measurement design that uses multiple traits (e.g., SDQ subscales) and multiple methods (e.g., parent and teacher reports). Convergent validity is demonstrated when different methods show agreement (i.e., a high correlation) for the same trait. Discriminant validity requires that conceptually different traits not be too highly correlated (Campbell & Fiske, 1959; Messick, 1995). Using the original MTMM approach based on observed scores, Goodman, Lamping, and Ploubidis (2010) found moderate support for convergent validity of the SDQ scales across reports from parents, teachers, and children aged 11 to 16 years. Similar levels of convergent validity and method effects were reported by Hill and Hughes (2007) using parent, teacher, and peer ratings of first-grade children at risk for educational failure. Both studies suggested poor discriminant validity, particularly for the CP, Hy, and PB subscales (Goodman et al., 2010; Hill & Hughes, 2007).

The use of observed scores that are not corrected for measurement error is a major limitation of the original MTMM approach. Measurement error can bias the correlation matrix, weakening the conclusions that can be drawn about convergent and discriminant validity. CFA-based techniques use latent variables that are free of measurement error and allow the separation of trait, method, and measurement error variance components (Gomez, 2014; Hill & Hughes, 2007; Van Roy, Veenstra, & Clench-Aas, 2008; Yu, Sun, & Cheah, 2016). Hill and Hughes (2007) fitted the correlated trait-correlated uniqueness (CT-CU) model (Kenny, 1976) to parent, teacher, and peer ratings, and van Roy et al. (2008) used the correlated trait-correlated method (CT-CM) model (Jöreskog, 1971; Widaman, 1985) to examine the SDQ's convergent validity and discriminant between parent ratings and preadolescent self-ratings. Overall, their results suggest poor discriminant validity, modest convergence across raters, and considerable method effects (Hill & Hughes, 2007; Van Roy et al., 2008).

Although the CT-CU and CT-CM models are among the most popular CFA models for analyzing MTMM data, several researchers have pointed out important shortcomings concerning their estimation and interpretation. The CT-CU model confounds true method effects and random measurement error. As a consequence, method variance cannot be isolated and indicator reliabilities are systematically underestimated by the model (Eid, Lischetzke, Nussbeck, & Trierweiler, 2003; Lance, Noble, & Scullen, 2002). The CT-CM model is not globally identified, frequently provides improper parameter estimates (e.g., negative variances), and can lead to results that are difficult to interpret (Eid et al., 2003; Fan & Lance, 2017).

To overcome issues with the CT-CU and CT-CM models, Eid and colleagues (Eid, 2000; Eid et al., 2003; Geiser, Eid, West, Lischetzke, & Nussbeck, 2012) proposed the correlated trait-correlated method minus one, CT-C(M − 1), model. The CT-C(M − 1) model avoids identification, estimation, and interpretation problems by defining trait and method factors based on classical psychometric test theory and including one method factor less than methods considered in the study. This is done by specifying one method as reference against which the remaining methods are contrasted in a latent regression analysis. The CT-C(M − 1) model allows determining the degree of convergent validity (consistency) between the reference method and all nonreference methods as well as the unique overlap between nonreference methods. In addition, the degree of method-specificity and reliability can also be estimated.

Two recent studies have used the CT-C(M − 1) model to investigate convergent and discriminant validity in the SDQ. Gomez (2014) analyzed data from adolescents collected through self-, mother, and teacher report, whereas Yu, Sun, and Cheah (2016) used data from preschool-aged children collected through mother and teacher reports. Both studies support discriminant validity between all trait factors measured by the SDQ, except between CP and Hy (Gomez, 2014; Yu et al., 2016). Findings from Gomez (2014) suggested high consistency between mothers' and adolescents' ratings, specifically for ES and PP. Teacher ratings showed modest consistency and a large proportion of their variance for CP, Hy, and PS was due to method specificity. Yu et al. (2016) also provided support for the convergence of mother and teacher ratings. However, their results suggested limited reliability of the ES subscale as reported by teachers.

Although the studies conducted by Gomez (2014) and Yu et al. (2016) yielded promising results regarding the SDQ's validity, their analyses were based on just a single indicator for each SDQ subscale. When using just a single indicator per subscale, rather restrictive assumptions about method effects must be made. That is, the single-indicator CT-C(M − 1) model implies that method effects for a given method (e.g., teacher reports) are perfectly homogeneous (perfectly correlated) across different traits. This assumption is frequently violated in practice because method effects tend to be partially trait-specific (Eid et al., 2003, 2008). Using single-indicator models when method effects are trait-specific can lead to bias in the estimated consistency, method-specificity, and/or reliability estimates. The use of multiple indicators within each trait-method combination is theoretically and empirically more appropriate as it allows specifying method factors that are specific to each subscale or trait (Eid et al., 2003, 2008).

Furthermore, previous studies have examined the convergent and discriminant validity of the SDQ using mother ratings as a proxy for parent ratings. No previous study has considered both mother and father ratings to examine the validity of the SDQ using MTMM models. Although some studies focusing on the SDQ's interparent agreement suggested high correlations between mother

and father reports (Davé et al., 2008; Mellor, Wong, & Xu, 2011), further evidence is needed to understand the degree to which mothers and fathers share a common perspective regarding their children's behavior.

The Current Study

The present study primarily focuses on the convergent and discriminant validity of the SDQ scales for ratings provided by teachers, mothers, and fathers. We start by examining the associations among the scores from these different informants and by analyzing whether teacher, mother, and father ratings differentially associate with children's characteristics, such as age, sex, and self-regulation abilities as directly assessed in a set of standardized tasks. Furthermore, we examined the SDQ's hypothesized five-factor structure and its level of MI across three raters, namely teachers, mothers, and fathers. Finally, we addressed convergent and discriminant validity by testing a multiple-indicator version of CT-C(M − 1) model with indicator-specific trait and method factors (Eid et al., 2008; the specific model is described in detail in the Method section).

Researchers frequently rely on parents and teachers to measure young children's behavior as expressed in the home and school settings. Although parents represent an important source of information regarding children's adjustment problems, teacher reports are frequently easier and less expensive to collect. In addition, teachers' perspectives can be particularly important and informative as they usually have an advanced training in education, regular experience with a large group of children, and the opportunity to repeatedly observe the child in a diverse set of situations and social interactions.

In the current study, the choice of reference method was made to provide the most theoretically convenient interpretation of the results. Therefore, we used teacher ratings as the reference method to more clearly contrast mother reports and father reports against this reference. This option allowed us to estimate the degree of consistency and method-specificity of mother and father ratings relative to teacher ratings. In addition, the tested model provided estimates of the level of convergence between mothers and fathers while controlling for their shared perspective with teachers. Based on previous research, we anticipated moderate to high levels of discriminant validity between the trait factors measured by the SDQ and moderate consistency between mother and teacher ratings (Gomez, 2014; Yu et al., 2016). We also expected fairly high levels of consistency between mother and father ratings, given that these informants tend to observe the child in similar contexts and situations (Davé et al., 2008).

Method

Participants

The participants in this study were 346 children (46% girls, *M* age = 54.07 months, *SD* age = 10.67). All children came from families with dual-earner and cohabiting parents and were attending private (52.73%) or public (48.27%) preschool centers in the metropolitan area of Porto, Portugal. The mothers' age ranged from 23 to 49 years (*M* = 35.55, *SD* = 4.57) and fathers' age ranged from 24 to 54 years (*M* = 36.94, *SD* = 5.14). A large percentage of parents completed higher education (62% of mothers, 43% of fathers). The teachers were all women with a university degree in education, aged between 22 and 54 years (*M* = 39.82, *SD* = 9.05). On average, each teacher reported on six children in their class.

Measures

The SDQ is a brief screening questionnaire used to evaluate the emotional and behavioral problems of children and adolescents, aged 4 to 16 years. Several versions are available, namely self-rated, parent-rated, and teacher-rated versions. Parent- and teacher-rated versions have the same 25 items arranged in five scales: ES (e.g., "Nervous or clingy in new situations, easily loses confidence"); PP (e.g., "Rather solitary, prefers to play alone"); CP (e.g., "Often fights with other children or bullies them"); Hy (e.g., "Easily distracted, concentration wanders"); and PB (e.g., "Helpful if someone is hurt, upset or feeling ill"). Each scale consists of five items, scored according to a 3-point scale ranging from 0 (*not true*) to 2 (*certainly true*).

In this study, we used the parent-rated version to collect both, mother and father reports, as well as the teacher-rated version to collect preschool teachers' reports on children's behavior. Preliminary analyses with this sample indicated a moderate internal consistency of mothers' reports, with a median Cronbach's alpha of .64, ranging from .41 to .70. Fathers' ratings had a median Cronbach's alpha of .65, ranging from .52 to .69. Teachers' ratings showed the highest internal consistency, with a median Cronbach's alpha of .71, ranging from .61 to .79. These reliability coefficients are similar to those found in previous research (Kersten et al., 2016; Stone et al., 2010) reflecting some internal consistency issues, particularly for parents scores. The diversity of behavioral aspects covered by the relatively small number of items in each scale might explain the moderate internal consistency coefficients. Besides, Stone, Otten, Engels, Vermulst, and Janssens (2010) suggested that teachers' ratings were more prone to a halo effect than parents' ratings, which might explain the differences in internal consistency between parents' and teachers' ratings. According to this explanation, items from the SDQ's subscales were less related to each other for parents than for teacher ratings due to higher discriminant validity of parent reports.

The Preschool Self-Regulation Assessment (PSRA; Cadima et al., 2016; Smith-Donald, Raver, Hayes, & Richardson, 2007) is a structured battery of tasks designed to evaluate children's self-regulation in emotional, attentional, and behavioral domains. Three tasks from the PSRA were selected and administered by trained and certified assessors, namely the Toy Sort, Toy Wrap, and Snack Delay. The Toy Sort task was designed to assess compliance with directions (Smith-Donald et al., 2007) and requires children to sort toys into boxes without playing with them. To score this task, the assessor records the time the child takes until completing the organizing task, without playing with the toys. The Toy Wrap and Snack Delay tasks tap children's ability to suppress a dominant response and undertake a subdominant response (Smith-Donald et al., 2007). In the Gift Wrap task, the child is told not to look to the "surprise" that the assessor is noisily wrapping. The score is obtained by recording the time until the child's first peek. Finally, the Snack Delay task requires children to wait before getting candy from under a cup. The assessor scores the

child's behavior in four different trials (10, 20, 30, and 60 s) using a 4-point rating, ranging from *does not touch cup* to *eats candy*. The task score is achieved by averaging the scores in all trials.

Procedure

The data used in this study was part of the baseline assessment of a broader longitudinal study aiming to understand the impact of work–family dynamics on parenting and child development. This research project was approved by the faculty's ethics committee and the schools' board. We recruited the teachers and families at the beginning of the school year. First, we explained the study to the teachers, who then invited the families of all the children in their classroom to participate. Parents' participation rate was 38%. The low participation rate was probably due to the characteristics of the recruitment process and the narrow eligibility requirements. The research team was only allowed to contact and collect information from families who agree to participate and signed the informed consent. The research team was only allowed to contact and collect information from families who agree to participate and signed the informed consent. After their written informed consent, teachers and parents were asked to fill in an individual questionnaire focusing on their parenting/teaching experience and on several indicators of the child's development, including the SDQ. Although there were no missing data in parents' reports, 29 children were not evaluated by their teachers, corresponding to 8.38% of all observations in teachers' data.

Data Analyses

Descriptive statistics were examined using composite scores for each SDQ scale as reported by teachers, mothers, and fathers. These scores were obtained by computing the mean of the item scores for each scale. Based on the SDQ's scales observed scores, we computed the zero-order correlations between teacher, mother, and father reports, investigating the same-trait, different-method (convergent validity), and the different-trait, same-method (discriminant validity) correlations. We also used zero-order correlations based on the SDQ observed scores to investigate the associations between teacher, mother, and father ratings with children's characteristics, namely sex, age, and self-regulation. Children's self-regulation was directly assessed by using a brief, structured battery of tasks. We adopted children's scores in these performance tasks as a criterion measure for establishing the predictive validity of teachers, mothers, and fathers SDQ's ratings.

Through structural equation modeling, we examined the SDQ's factorial structure and investigated its validity and method specificity. We conducted CFA-MTMM analyses to examine the convergent and discriminant validity of mothers', fathers', and teachers' SDQ ratings. Following Geiser and colleagues (Geiser et al., 2014, 2012; Geiser, Mandelman, Tan, & Grigorenko, 2016), the analytical plan proceeded in three main steps. First, the SDQ's factor structure was evaluated through separate CFA models. A representation of the CFA model is displayed in Figure 1a. This model specifies five latent factors representing ES, PP, CP, Hy, and PB for each type of rater—teacher, mothers, and fathers. The latent factors for the distinct behavioral dimensions can be interpreted as rater-specific common true score variables. All latent factors were freely correlated. Estimated correlations between rater-specific factors pertaining to the same trait but a different rater type are indicative of interrater agreement (convergent validity). Estimated correlations between rater-specific factors pertaining to the different traits are indicative of discriminant validity. In the present study, we examined convergent and discriminant validity in more detail using the CT-C(M − 1) approach as explained below.

Second, we tested for MI by comparing CFA models representing weak, strong, and strict MI across rater types. CFA models with different levels of MI are nested models and can thus be directly compared using chi-square difference tests (Satorra & Bentler, 2001). We also computed the differences in comparative fit index (CFI) values to examine the magnitude of differences between models (Cheung & Rensvold, 2002).

Third, an adaptation of the CT-C(M − 1) model was tested to examine the convergent validity and method specificity of mother, father, and teacher ratings in more detail. A representation of the tested model is displayed in Figure 1b. The CT-C(M − 1) model includes two types of latent variables: (a) the reference factor, representing the trait as measured by a selected reference method (here: teacher reports); and (b) the method factors, which represent residual variance in the nonreference methods (here: mother and father reports) that is not shared with the reference factor pertaining to the same trait. In addition, measurement error variables are included to separate true trait and true method variance from variance due to random measurement error.

As recommended by Eid and colleagues (Eid et al., 2003, 2008), the CT-C(M − 1) model tested in this study (Figure 1b) used multiple indicators for each trait and indicator-specific trait factors, including three indicators per trait-method-unit (TMU). The model includes five traits and three methods with indicator-specific trait factors. This indicator-specific or item-level approach differs from a trait-level approach which specifies a single general trait factor, assuming that the multiple indicators are perfectly homogeneous representations of the construct. By including indicator-specific or item-level trait factors, the model in the present study accounts for the fact that indicators of the same TMU may refer to slightly different facets of a trait or construct (Eid et al., 2008). Because there are three indicators for each trait-method combination (see Figure 1b) there are three latent trait variables for each construct.

In the CT-C(M − 1) model, nonreference methods are contrasted against the reference factor. Given that method factors are defined as regression residuals with respect to the reference factors in the model, reference and method factors for the same trait are uncorrelated. Therefore, two independent variance components (reference trait vs. method) and related coefficients can be computed. The *consistency coefficient* gives the proportion of variance of an indicator that is explained by the reference method factor and can be interpreted as an index of convergent validity. The *method specificity coefficient* gives the proportion of the variance of a nonreference indicator explained by a nonreference method factor, representing the degree of method specificity. Consistency and method specificity coefficients for the observed indicators can be obtained based on the squared standardized factor loadings (SSLs). The sum of consistency plus method specificity for a given indicator provides an estimate of that indicator score's reliability (see Eid et al., 2003, 2008; as well as Geiser et al., 2016; for further details on the CT-C(M − 1) model).

We used teachers' ratings as the reference method and contrasted against mothers' and fathers' ratings. This allowed us to quantify the extent to which mothers' and fathers' ratings converged with teach-

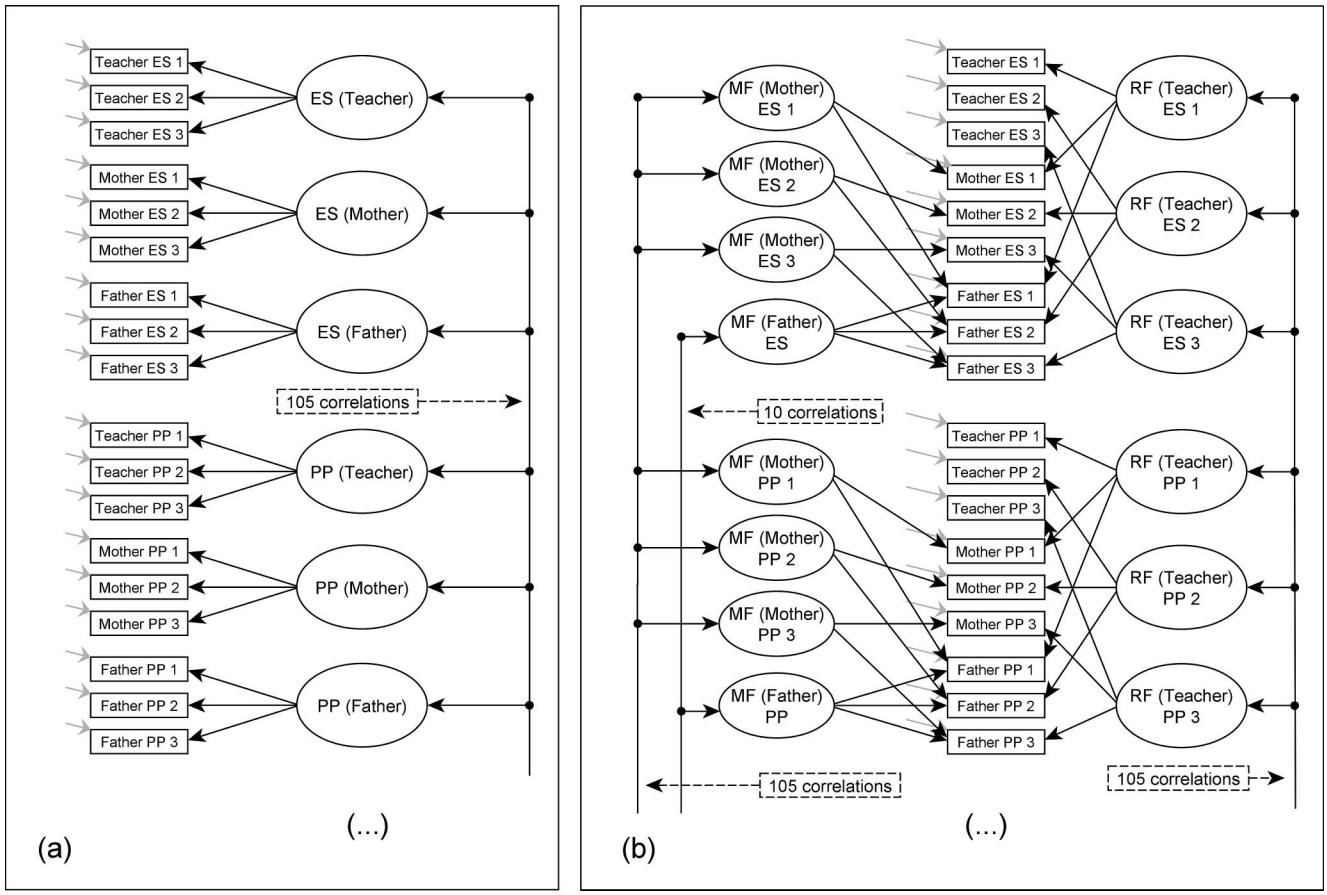

Figure 1. Simplified representation of (a) CFA model and (b) CT–C(M − 1) models. RF = reference factor; MF = method factor; ES = emotional symptoms; PP = peer problems.

ers' ratings. In addition, mothers' ratings were specified as a reference method for fathers' ratings to examine the extent to which mothers' and fathers' unique views of the children (relative to the teachers) overlapped. This adaptation to the original way of specifying the CT-C(M − 1) model allowed us to capture the level of convergence between mothers and fathers while controlling for their shared perspective with teachers.

We used parcels of items to define the latent variables representing ES, PP, CP, Hy, and PB measured through mother, father, and teacher ratings. This option offers several advantages, reducing the number of parameters in the models and contributing to less biased estimates (Coffman & MacCallum, 2005). Each factor was defined by one observed item and two parcels. Parcels were computed using two randomly selected items. We used the same parcels of items for mother, father, and teacher corresponding factors. All models—that is, the CFA, MI, and CT-C(M − 1) models—were refitted using four different items-to-parcel random combinations to verify whether item-to-parcel allocation would alter any of the study conclusions. The results matched the findings reported throughout this paper, suggesting negligible consequences of item-to-parcel allocation.

Following Widaman and Thompson (2003), we used the appropriate null model for multirater comparisons in our computation of relative fit indices (CFI and Tucker-Lewis index [TLI]). All models were identified through the marker variable method, fixing the loading and intercept of the first indicator of each construct to one and zero, respectively. Model fit was examined using the chi-square goodness-of-fit statistic, the root mean square error of approximation (RMSEA), the comparative fix index (CFI), the TLI, and the standardized root-mean-square residual (SRMR). Values lower than .06 for RMSEA, greater than .95 for CFI and TLI, and lower than .08 for SRMR indicate good model fit (Hu & Bentler, 1999).

All analyses were conducted in R (R Core Team, 2018), using the "mice" (Van Buuren & Groothuis-Oudshoorn, 2011), the "lavaan" (Rosseel, 2012), and the "lavaan.survey" packages (Oberski, 2014). Missing data were imputed through multiple imputations by chained equations (10 imputations). Imputation was conducted at the parcel level. All models were estimated using maximum likelihood estimation with robust standard errors.[1] We adopted a design-based approach to account for the fact that children were nested within classrooms (Muthén & Satorra, 1995). In this ap-

[1] To inspect whether using an estimator for categorical data would return different results than the ones obtained with maximum likelihood estimation with robust standard errors, we re-tested all models using a robust weighted least square estimation method (WLSMV). We used Mplus to conduct these analyses (Muthén & Muthén, 2010). The two analytical approaches yielded comparable results (see the online supplemental material for a detailed description of the results using WLSMV estimation).

proach, parameter estimates are aggregated according to a cluster variable (classroom) and standard errors are corrected for potential nonindependence of observations.

Results

Descriptive Statistics

Means, standard deviations, and zero-order observed correlations for the observed variables can be found in Table 1. Teachers, mothers, and fathers reported low average levels of ES, PP, CP, and Hy, and high average levels of PB. For the observed variables, the average same-trait, different-method (convergent validity) correlation between teacher, mother, and father reports was .35, ranging from .23 to .58. The average absolute different-trait, same-method (discriminant validity within the same method) correlation was .26, ranging from .03 to .56.

Table 1 also depicts the correlations of teacher, mother, and father SDQ's ratings with children's age, sex, and self-regulation scores on the PSRA selected tasks, namely Toy Sort, Toy Wrap, and Snack Delay. Children's age was positively associated with teacher, $r = .19$, $p < .001$; mother, $r = .13$, $p = .012$; and father, $r = .16$, $p = .003$ ratings of PB. Boys tended to display higher levels of CP, $r = .25$, $p < .001$, and Hy, $r = .17$, $p = .002$, as well as lower levels of PB, $r = -.19$, $p < .001$, as reported by teachers. A larger number of significant correlations emerged between children's self-regulation and teachers' SDQ ratings than between self-regulation and mother and father ratings. Differences in the pattern of correlations were particularly evident in the Toy Wrap and Snack Delay tasks, which specifically tap children's ability to suppress a dominant response and undertake a subdominant response. Children's performance on the Toy Wrap task was negatively associated with teachers' rating on CP, $r = .19$, $p = .008$, and Hy, $r = .18$, $p = .010$; and positively associated with teachers reported PB, $r = .16$, $p = .030$. Also, there were significant associations between children's performance on the snack delay task significantly and teachers' ratings across all the SDQ subscales.

Confirmatory Factor Analysis and Measurement Invariance

The SDQ's five-factor CFA solution with three indicators (parcels) for each trait was first fit to teachers, mothers, and fathers data separately. As shown in Table 2, this CFA model provided a good overall fit to the data from teachers: $\chi^2(80) = 101.61$, $p = .052$, RMSEA = .03 (90% CI [.01, .04]), CFI = .98, TLI = .97, SRMR = .05; and fathers: $\chi^2(80) = 124.85$, $p = .001$, RMSEA = .04 (90% CI [.03, .06]), CFI = .96, TLI = .94, SRMR = .05; and marginal acceptable fit to mothers' data: $\chi^2(80) = 151.60$, $p < .001$, RMSEA = .06 (90% CI [.04, .07]), CFI = .91, TLI = .88, SRMR = .06.

The tests of MI across raters were conducted by imposing sequential equality constraints on loadings, intercepts, and residual variances (see Table 2). The baseline model for testing MI included 15 latent variables representing each trait measured by teachers, mothers, and fathers. The residual variances of the corresponding indicators measured by the different raters were allowed to correlate. The configural invariance model (Model A) fit the data adequately, $\chi^2(795) = 983.91$, $p < .001$, RMSEA = .03 (90% CI [.02, .03]), CFI = .96, TLI, = .95, SRMR = .05. Results from the chi-square difference test and CFI difference indicated that the weak MI model (Model B) did not fit significantly worse than the configural MI model (Model A), $\Delta\chi^2(20) = 24.97$, $p = .203$, ΔCFI = 0.001, thereby establishing metric invariance (Cheung & Rensvold, 2002). The assumption of strong MI across all three rater types did not hold as indicated by the significant decline in model fit when comparing the Model B to the strong MI model (Model C), $\Delta\chi^2(20) = 324.73$, $p < .001$, ΔCFI = 0.028.

We subsequently tested a partial MI model (Model D) by removing the imposed equality constraints on five teacher report intercepts, one for each trait factor. These five item intercepts were unconstrained considering that their values were the ones that most significantly deviated from the values of the intercepts for the corresponding items from mother and fathers reports. Specifically, we freed the teacher intercepts pertaining to the second parcel for the ES: Item 8 ("worries") and Item 13 ("unhappy"); PP: Item 11 ("good friend") and Item 19 ("bullied"); CP: Item 5 ("tempers") and Item 12 ("fights"); PB: Item 4 ("shares") and Item 9 ("caring") subscales as well as the third parcel for the Hy: Item 10 ("fidgety") and Item 21 ("thinks before acting") subscale.

Globally, the noninvariant item intercepts for teacher ratings were higher than the item intercepts for mother and father ratings, indicating that teachers more easily endorse these parcels of items than parents. This may have been the case because most of the noninvariant parcels included items tapping into behavior problems that more likely emerge in social interactions, namely with other children; for example, Item 19 ("bullied"), Item 12 ("fights"), Item 4 ("shares"), and Item 10 ("fidgety"). These social interactions frequently take place in the school context and can more easily be observed by teachers than parents.

Although the chi-square difference test was significant, $\Delta\chi^2(15) = 79.01$, $p < .001$, the change of CFI between Model D and the weak MI model (Model B) was negligible (Cheung & Rensvold, 2002), ΔCFI = 0.009, indicating that strong MI could be assumed across mother and father ratings and that partial strong MI could be assumed across parent (mother and father) and teachers ratings.

The partial strong MI model was retained and used to test for strict MI across mother and father ratings only, leaving the residual variance parameters for teachers unconstrained. The decrease in model fit from Model D to the partial strict MI model (Model E) was not statistically significant, $\Delta\chi^2(15) = 20.63$, $p = .149$, ΔCFI = 0.002, indicating that the residual variances did not differ across mother and father ratings.

In summary, the SDQ scales satisfied the condition of strict MI across mother and father reports whereas teacher reports showed equivalent loadings, partially equivalent intercepts, but nonequivalent residual variances relative to parent reports. This level of MI allowed for a direct comparison of latent means across teachers, mothers, and, fathers (Byrne, Shavelson, & Muthén, 1989).

Table 3 includes the factor loadings and items intercepts in the measurement model with partial strict invariance. All indicators showed moderate to high standardized factor loadings on the trait factors (range = .37; .88).

Table 4 displays the latent means, variances, and correlations estimated in the CFA model with partial strict MI. The latent means were lower for teacher ratings than for mother and father ratings across all subscales. Compared with mothers, fathers re-

Table 1
Means, Standard Deviations, and Zero-Order Correlations for the Observed Variables

		Teacher					Mother					Father				
Indicator	*M* (*SD*)	ES	PP	CP	Hy	PB	ES	PP	CP	Hy	PB	ES	PP	CP	Hy	PB
Teacher																
ES	0.29 (0.33)	—														
PP	0.16 (0.16)	.38**	—													
CP	0.27 (0.27)	.03	.17**	—												
Hy	0.57 (0.57)	.04	.17**	.54**	—											
PB	1.59 (0.42)	−.19**	−.41**	−.32**	−.34**	—										
Mother																
ES	0.40 (0.33)	.28**	.18**	−.10	−.03	−.05	—									
PP	0.27 (0.27	.14*	.23**	−.03	.02	−.16**	.37**	—								
CP	0.56 (0.37)	.07	.03	.30**	.23**	−.13*	.15**	.20**	—							
Hy	0.82 (0.46)	.02	.04	.27**	.44**	−.15*	.12*	.13**	.52**	—						
PB	1.66 (0.31)	.01	−.07	−.15*	−.14*	.24**	−.05	−.19**	−.27**	−.20**	—					
Father																
ES	0.44 (0.35)	.26**	.20**	−.03	−.07	.00	.47**	.23**	.06	−.01	−.08	—				
PP	0.31 (0.30)	.17**	.23**	−.03	−.08	−.09	.21**	.40**	.08	−.03	−.14**	.48**	—			
CP	0.55 (0.37)	−.01	.05	.28**	.17**	−.10	−.01	.06	.54**	.33**	−.27**	.23**	.25**	—		
Hy	0.86 (0.45)	−.06	.06	.33**	.33**	−.11*	−.01	.04	.42**	.58**	−.18**	.16**	.13**	.56**	—	
PB	1.62 (0.33)	−.05	−.14*	−.10	−.03	.23**	−.05	−.19**	−.20**	−.13**	.39**	−.19**	−.40**	−.35**	−.28**	—
Child																
Age	54.07 (10.67)	−.18**	−.04	−.04	−.05	.19**	.00	−.10	−.14*	−.03	.13*	−.05	−.19**	−.09	.01	.16**
Sex (1 = boy)	0.54 (0.50)	.01	−.03	.25**	.17**	−.19**	−.04	−.02	.09	.16**	−.06	−.07	−.03	.07	.10	−.08
Toy Sort	80.55 (31.61)	.14*	.14	−.06	.00	−.09	.19*	.21**	.19*	.10	−.10	.16*	.27**	.12	.12	−.16*
Toy Wrap	90.28 (45.03)	−.06	−.08	−.19*	−.18*	.16*	−.02	.00	−.11	−.05	.05	.02	−.08	−.04	−.07	.10
Snack Delay	3.69 (0.52)	−.21**	−.27**	−.24**	−.27**	.29**	−.03	−.11	−.18*	−.18*	.07	−.06	−.12	−.12	−.09	.16*

Note. ES = emotional symptoms; PP = peer problems; CP = conduct problems; Hy = hyperactivity; PB = prosocial behaviors.
* $p < .05$. ** $p < .01$.

Table 2
Model Fit Information for Different CFA Models

Model tested	Goodness of fit					Model comparison		
	$\chi^2(df)$	RMSEA [90% CI]	CFI	TLI	SRMR	Compared model	$\Delta\chi^2(\Delta df)$	ΔCFI
CFA model								
Teacher report	101.61 (80)	.03 [.01, .04]	.98	.97	.05	—	—	—
Mother report	151.60 (80)**	.06 [.04, .07]	.91	.88	.06	—	—	—
Father report	124.85 (80)*	.04 [.03, .06]	.96	.94	.05	—	—	—
MI model								
Null model	5538.84 (1050)**	—	—		—	—	—	—
Configural MI (A)	983.91 (795)**	.03 [.02, .03]	.96	.95	.05	—	—	—
Metric MI (B)	1008.93 (815)**	.03 [.02, .03]	.96	.94	.05	A	24.97 (20)	.001
Strong MI (C)	1156.72 (835)**	.04 [.03, .04]	.93	.91	.06	B	324.73 (20)**	.028
Partial strong MI (D)	1067.36 (830)**	.03 [.03, .04]	.95	.93	.05	B	79.01 (15)**	.009
Partial strict MI (E)	1087.88 (845)**	.03 [.03, .04]	.95	.93	.06	D	20.63 (15)	.002

Note. CFA = confirmatory factor analysis; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root-mean-square residual; MI = measurement invariance.
* $p < .05$. ** $p < .01$.

ported slightly higher average levels of ES, PP, and Hy and lower average levels of CP and PB. Wald tests revealed that the latent mean differences were significant between mother, father, and teachers for all of the SDQ subscales, except for CP reported by mothers and fathers, $\chi^2(1) = 0.17$, $p = .683$, and for PB reported by teachers and fathers, $\chi^2(1) = 3.07$ $p = .080$.

The average latent correlation between teacher, mother, and father reports of the same trait (convergent validity correlations) was .45, ranging from .31 to .74. The latent correlations between the same latent factors reported by mothers and fathers ranged from .48 to .74, with an average correlation of .60, whereas the correlations between parent (mother and father) and teacher reports of the same trait ranged from .31 to .53, averaging .37. The average absolute correlation between different traits reported by the same rater type (discriminant validity correlation) was .37 (range = .03; .84).

Table 3
CFA (Partial Strict MI Model) Factor Loadings and Item's Intercepts

Indicator	Factor loadings (US/S)			Item intercepts (US)		
	Parcel 1[a]	Parcel 2	Parcel 3	Parcel 1[a]	Parcel 2	Parcel 3
Teacher						
ES	1/.60	0.57/.53	0.88/.69	0	.01	.11
PP	1/.61	0.57/.68	0.83/.72	0	.03	.05
CP	1/.69	0.82/.72	0.61/.67	0	.07	.07
Hy	1/.48	1.38/.88	1.12/.79	0	−.10	−.01
PB	1/.67	0.95/.83	0.79/.72	0	.07	.44
Mother						
ES	1/.50	0.57/.54	0.88/.62	0	−.13	.11
PP	1/.45	0.57/.37	0.83/.49	0	.08	.05
CP	1/.61	0.82/.77	0.61/.63	0	−.12	.07
Hy	1/.50	1.38/.83	1.12/.70	0	−.10	−.08
PB	1/.57	0.95/.64	0.79/.62	0	−.05	.44
Father						
ES	1/.54	0.57/.58	0.88/.65	0	−.13	.11
PP	1/.54	0.57/.45	0.83/.58	0	.08	.05
CP	1/.59	0.82/.76	0.61/.62	0	−.12	.07
Hy	1/.48	1.38/.82	1.12/.68	0	−.10	−.08
PB	1/.58	0.95/.66	0.79/.64	0	−.05	.44

Note. ES = emotional symptoms; PP = peer problems; CP = conduct problems; Hy = hyperactivity; PB = prosocial behaviors; US = unstandardized estimates; S = standardized estimates; CFA = confirmatory factor analysis; MI = measurement invariance.
[a] Parameter fixed for identification.

CT-C(M − 1) Model

The use of the CT-C(M − 1) model allowed us to investigate the consistency and method-specificity in more detail. Teacher reports were selected as reference method, against which both mother and father reports were contrasted. In addition, mother method factors were defined as reference factors relative to father reports, allowing the examination of their level of convergence above and beyond what both mother and father reports had in common with teacher reports. The model structure is displayed in Figure 1b.

The CT-C(M − 1) model fit the data adequately, $\chi^2(680) = 751.21$, $p = .030$, RMSEA = .02 (90% CI [.01, .03]), CFI = .98, TLI = .97; SRMR = .04. Table 5 shows the estimated factor loadings for the indicator-specific reference and method factors. The standardized teacher reference trait factor loadings for the parcels pertaining to mother and father reports were fairly low, ranging from .13 to .43. This reflected the rather modest convergence of parents and teacher reports. The loadings of mother and father reports on their respective method factors were larger than the reference trait loadings. Taken together, this indicated that there was higher method specificity than convergent validity of both parents' reports relative to teacher reports. Father report parcels had similar loadings on the mother method factor (range = .30; .61) and on the method factor unique to father ratings (range = .39; .62). This indicated that there was both some shared variance across mothers and fathers as well as some unique father report variance that was not shared with mothers.

Consistency, method specificity, and reliability coefficients are presented in Table 6. The consistency coefficients (i.e., the proportion of the observed variance in a nonreference indicator that was shared with the reference method) were very low for all traits (range = .02; .18). Overall, the teacher reference trait factors explained less than 10% of the variance in the majority of the indicators pertaining to mothers and fathers, once again indicating

Table 4
CFA (Partial Strict MI Model) Latent Means, Variances, and Correlations

			Correlations														
			Teacher					Mother					Father				
Indicator	*M* (*SE*)	Variance (*SE*)	ES	PP	CP	Hy	PB	ES	PP	CP	Hy	PB	ES	PP	CP	Hy	PB
Teacher																	
ES	0.32 (.04)	.11 (.02)	—														
PP	0.18 (.03)	.09 (.03)	.53	—													
CP	0.27 (.04)	.16 (.02)	.03	.24	—												
Hy	0.50 (.04)	.11 (.01)	.04	.19	.71	—											
PB	1.54 (.05)	.17 (.03)	−.27	−.55	−.43	−.40	—										
Mother																	
ES	0.52 (.03)	.10 (.02)	.36	.27	−.15	−.05	−.06	—									
PP	0.27 (.03)	.05 (.01)	.30	.34	−.07	−.01	−.28	.72	—								
CP	0.76 (.04)	.17 (.02)	.12	.07	.39	.30	−.18	.23	.32	—							
Hy	0.75 (.03)	.11 (.01)	.03	.06	.37	.53	−.18	.11	.16	.69	—						
PB	1.67 (.03)	.08 (.01)	.01	−.13	−.22	−.18	.33	−.08	−.33	−.35	−.27	—					
Father																	
ES	0.58 (.03)	.13 (.03)	.35	0.29	−.04	−.10	.01	.62	.49	.09	−.02	−.15	—				
PP	0.34 (.03)	.08 (.02)	0.33	.31	−.06	−.13	−.11	.42	.51	.12	−.04	−.20	.84	—			
CP	0.75 (.04)	.16 (.02)	.02	.08	.37	.21	−.12	.01	.04	.65	.45	−.36	.36	.42	—		
Hy	0.79 (.03)	.10 (.01)	−.06	.08	.45	.43	−.15	−.09	.02	.55	.74	−.24	.16	.16	.76	—	
PB	1.62 (.02)	.09 (.02)	−.07	−.23	−.14	−.04	.32	−.11	−.31	−.31	−.22	.48	−.29	−.58	−.54	−.41	—

Note. ES = emotional symptoms; PP = peer problems; CP = conduct problems; Hy = hyperactivity; PB = prosocial behaviors; *SE* = standard error; CFA = confirmatory factor analysis; MI = measurement invariance. All correlation above .15 and below −.15 were statistically significant at the .05 level.

that both parent reports showed low convergent validity with respect to the teacher report.

The method-specificity coefficients represent the reliable proportion of observed variance that is specific to a particular nonreference method. The CT-C(M − 1) model allowed us to estimate two method-specificity coefficients: (a) both parents' shared method-specificity relative to teacher reports and (b) fathers' unique method-specificity relative to mother reports. Parents' method-specificity coefficients were high for most of the mother ratings (range = .33; .74), indicating that a large proportion of variance in mother reports was not shared with teacher reports. On average, father ratings shared 20% (range = 9%; 38%) of their variance with mother ratings. On the other hand, on average, 25% (range = 15%; 38%) of the variance in father ratings was unique to fathers in that it was neither shared with teachers nor with mothers.

Finally, the reliability coefficients represent the proportion of variance in each observed indicator that represented true score variance and was not due to measurement error. The reliability coefficients were moderate to high for all indicators, ranging from .35 to .88 (M = .59). The average reliability coefficient was higher for mother indicators (M = .61) than for father indicators (M = .51).

Table 7 presents the range of the latent correlations between the indicator-specific trait and method factors. Results indicated that there was some variability in the correlations involving indicator-specific factors pertaining to the same trait. A wide range of correlations between indicator-specific trait factors was found between PP and CP (range = .00; .43), ES and PB (range = −.41; −.06), and between CP and PB (range = −.48; −.13). The lowest range of correlations between indicator-specific trait factors was found between ES and PP (range = .30; .49), and between PP and PB (range = −.52; −.34). These correlations indicated that there was low to moderate overlap between indicator-specific traits representing different parcels. This means that the different parcels represented fairly distinct facets of the constructs studied here.

Table 5
Range of Factor Loadings Obtained in the CT-C(M - 1) Analysis

	Standardized reference factor loadings (Range)			Standardized method factor loadings[a] (Range)		Standardized method factor loadings[b] (Range)
Indicator	Teacher report[c]	Mother report	Father report	Mother report[c]	Father report	Father report
Emotional symptoms	[.74, .81]	[.24, .36]	[.18, .35]	[.68, .76]	[.32, .61]	[.39, .56]
Peer problems	[.72, .93]	[.13, .31]	[.17, .28]	[.65, .86]	[.30, .52]	[.43, .48]
Conduct problems	[.64, .79]	[.23, .34]	[.19, .27]	[.69, .77]	[.42, .49]	[.41, .62]
Hyperactivity	[.77, .89]	[.33, .43]	[.23, .38]	[.57, .74]	[.45, .54]	[.42, .57]
Prosocial behaviors	[.71, .94]	[.13, .25]	[.16, .26]	[.72, .81]	[.31, .39]	[.51, .58]

Note. CTC- C = correlated trait-correlated uniqueness. Model used teacher ratings as indicators of the reference factor. All factor loadings were significantly different from zero ($p < .05$).
[a] Method factor combining items from mother and father reports. [b] Method factor including items from father report. [c] Parameter fixed for identification.

Table 6
Variance Components in the CT-C(M - 1) Model With Teacher Reports as Reference Method

Trait and method	Consistency	Method specificity[a]	Method specificity[b]	Reliability
Emotional symptoms				
Teacher 1				.65
Teacher 2				.57
Teacher 3				.55
Mother 1	.13	.46		.59
Mother 2	.08	.57		.65
Mother 3	.06	.52		.58
Father 1	.12	.20	.21	.53
Father 2	.03	.10	.31	.45
Father 3	.07	.38	.15	.60
Peer problems				
Teacher 1				.71
Teacher 2				.52
Teacher 3				.87
Mother 1	.06	.42		.49
Mother 2	.02	.58		.59
Mother 3	.10	.74		.83
Father 1	.03	.27	.23	.53
Father 2	.03	.18	.22	.43
Father 3	.08	.09	.19	.35
Conduct problems				
Teacher 1				.41
Teacher 2				.63
Teacher 3				.56
Mother 1	.08	.47		.55
Mother 2	.11	.59		.70
Mother 3	.05	.49		.54
Father 1	.07	.18	.25	.49
Father 2	.03	.23	.38	.64
Father 3	.06	.24	.17	.47
Hyperactivity				
Teacher 1				.59
Teacher 2				.8
Teacher 3				.68
Mother 1	.11	.42		.53
Mother 2	.16	.55		.72
Mother 3	.18	.33		.51
Father 1	.05	.20	.32	.58
Father 2	.14	.29	.18	.61
Father 3	.08	.28	.21	.57
Prosocial behaviors				
Teacher 1				.63
Teacher 2				.88
Teacher 3				.51
Mother 1	.06	.52		.58
Mother 2	.02	.66		.68
Mother 3	.06	.55		.62
Father 1	.03	.15	.34	.52
Father 2	.03	.15	.27	.44
Father 3	.07	.10	.26	.42

Note. CTC-C = correlated trait-correlated uniqueness.
[a] Method factor combining items from mothers and fathers reports. [b] Method factor including items from fathers report.

There was also a wide range of correlations between the indicator-specific method factors defined by mother and father reports across different constructs. The highest range was observed for the correlation between ES and Hy (range = −.09; .44), and between CP and PB (range = −.59; −.05), while the lowest range was observed on the correlation between ES and CP (range = .03; .26), ES and PP (range = .19; .43), and between ES and PB, (range = −.18; .08). These correlations indicated a low to moderate generalization of method effects across constructs, underlining the importance of specifying trait-specific method factors.

Discussion

The current study addressed the validity of Strength and Difficulties Questionnaire (SDQ), a widely used measure for screening children's psychological adjustment. We examined the adequacy of the five-factor model originally proposed by Goodman (1997)

Table 7
Range of Latent Correlations Between Indicator-Specific Trait and Between Method Factors

Indicator	ES	PP	CP	Hy	PB
Trait					
ES	—				
PP	[.30; .49]	—			
CP	[−11; .15]	[.00; .43]	—		
Hy	[−.17; .23]	[.04; .38]	[.40; .72]	—	
PB	[−.41; −.06]	[−.52; −.34]	[−.48; −.13]	[−.43; −.24]	—
Method factor 1[a]					
ES	—				
PP	[.19; .43]	—			
CP	[.03; .26]	[.04; .37]	—		
Hy	[−.09; .44]	[−.08; .27]	[.31; .64]	—	
PB	[−.18; .08]	[−.37; .07]	[−.59; −.05]	[−.28; .04]	—
Method factor 2[b]					
ES	—				
PP	.82	—			
CP	.63	.58	—		
Hy	.51	.39	.76	—	
PB	−.30	−.58	−.48	−.40	—

Note. ES = emotional symptoms; PP = peer problems; CP = conduct problems; Hy = hyperactivity; PB = prosocial behaviors; US = unstandardized estimates; S = standardized estimates. All correlation above .19 and bellow −.19 were statistically significant at the .05 level.
[a] Method factor combining items from mothers and fathers reports. [b] Method factor including items from fathers report.

and its invariance across teacher, mother, and father reports on the same child. We also investigated the level of convergence between teacher, mother, and father ratings as well as the discriminant validity of the five traits measured by the SDQ.

Factor Structure and Measurement Invariance

Results indicated that the SDQ's original five-factor solution provides acceptable to good overall fit to the data collected through teacher, mother, and father reports, suggesting an equivalent underlying factor configuration across raters. In addition, MI analysis revealed that the SDQ's factor loadings were not significantly different across teacher, mother, and father reports. This finding is generally consistent with results from previous works (Rogge et al., 2018; Sanne et al., 2009), suggesting that teacher, mother, and father reports of SDQ share the same units of measurement.

In addition, our results supported the assumption of strong invariance across mother and father reports. This indicates that also the origins of measurement (item difficulties) are generally comparable across mother and father reporters. Strong MI is a required condition for meaningful comparisons of latent means across reporters. Therefore, this finding is important as it indicates that the SDQ may allow for a meaningful examination of convergent validity across mother and father ratings also with regard to the mean levels of child problem behavior.

Although partial strong invariance between teacher, mother, and father ratings was achieved, the number of noninvariant intercepts was notable, suggesting the need for further study. In line with previous studies (Geiser et al., 2014; Sanne et al., 2009), we found that, on average, teacher reports indicated a lower average level of behavior problems than did mother and father reports. Perhaps teachers observe a broader spectrum of behavior difficulties in class than do parents at home. This may lead to a different frame of reference for teachers resulting in lower average ratings. Also, teachers likely evaluate a given child relative to other children in the class, which may also lead to lower average rates of behavior problems. Another interpretation could be that the level of problem behavior is truly lower within the school as compared with the home context, for example, due to a more highly structured environment at school.

So far, only Rogge et al. (2018) provided support for a level of MI higher than metric, showing that strong and strict invariance between parents and teachers ratings is tenable. Unlike Rogge et al. (2018), the assumption of full strong MI between SDQ scores from mothers, fathers, and teachers did not hold in the current study. Nevertheless, our results supported the assumption of partial strong invariance across parents and teachers ratings. This level of MI still allows for meaningful comparisons of factor means across rater types.

To our knowledge, this is the first study examining the level of MI across mothers and fathers for the SDQ measure. Our results suggest that the latent constructs are measured in a strictly equivalent way across mother and father reports. The higher level of MI between mothers and fathers as compared with teachers likely reflects the influence of the context in which the children's behavior is observed. That is, mothers and fathers observe their children in the same context which may explain the higher level of MI.

Reliability, Convergent Validity, and Discriminant Validity

In addition to the invariance of the factor structure across different raters, our analyses with the CT-C(M − 1) model allowed us to investigate the SDQ's reliability as well as convergent and

discriminant validity in more detail by estimating the reliability, consistency, and method-specificity coefficients relative to a reference method (teacher reports).

The present study clarifies the extent to which mother and father ratings of children's behavior converged with the teachers' reports and the extent to which this convergence reflects true convergent validity rather than shared rater biases. Overall, the results suggest that both mother and father ratings can provide reliable measures of children's psychological adjustment. Measurement error variances did not differ significantly between mother and father reports indicating comparable measurement precision of corresponding mother and father report scales.

Mother and father reports showed lower levels of convergent validity relative to teacher's reports. The relatively low level of convergence between parents and teacher ratings was not surprising. Previous studies have reported similar results, arguing that the limited convergence between informants reflects real differences in children's behavior due to the context of observation (Gomez, 2014). Our results support the situational specificity hypothesis (Achenbach, McConaughy, & Howel, 1987), indicating that mothers' and fathers' low levels of convergent validity relative to teacher ratings is due to the fact that parents and teachers observe and interact with the child in distinct contexts (i.e., home vs. school). Accordingly, the degree of method specificity was notable, indicating the existence of a large proportion of observed variance in parents' ratings of the SDQ that, although reliable, is not shared with the teachers.

Despite the specificity of both parents' perspective relative to teachers, it should be noted that mothers and fathers may have different perspectives on children's behavior. The current study makes a noteworthy contribution by clarifying the proportion of variance in father reports of the SDQ that is not shared with mother reports. Our results suggest that fathers have a unique view on children's behavior that goes above and beyond the mother (and teacher) reports. This finding suggests that even though mother and father reports show relatively high agreement, they also have (partly) different perspectives about their child's behavior—despite the fact that they observe their children in similar contexts and situations.

According to De Los Reyes, Thomas, Goodman, and Kundey (2013), meaningful discrepancies between informants can result from systematic differences in three aspects: the behavior interpretation, the decision threshold for identifying the behavior, and the observation context. The modest convergence between parents and teacher reports of SDQ suggests that the observation context might be a major source of informant discrepancy. Parents and teacher reports might reflect an expected variation in the expression of child behavior across the home and school settings. On the other side, because mothers and fathers observe child behavior in similar settings, the discrepancy between their ratings is probably due to distinct interpretations of why the child is expressing the behavior under assessment and to differences in the decision thresholds underlying their judgment of child behavior.

The results concerning discriminant validity are consistent with previous studies (Gomez, 2014; Yu et al., 2016), indicating poor discriminant validity between CP and Hy. In addition, we also found substantial overlap between the PP and PB factors. As suggested by Gomez (2014), the poor discriminant validity between CP and Hy might be explained by the fact that these two dimensions are measuring a single higher-order factor representing externalizing problem behaviors. Indeed, some previous studies pointed out some advantages of using a second-order internalizing and externalizing factors, particularly for evaluating low-risk samples (Goodman et al., 2010), such as the one used in the current study. The existence of a second-order factor that includes the CP and Hy items may contribute to a poor discriminant validity between these two factors.

Similarly, the overlap between PP and PB may be explained by the fact that these two factors are mostly measured by indicators that clearly connect to the way children behave in social interactions with peers; for example, "generally liked by other children" (PP); "shares readily with other children, for example toys, treats, pencils" (PB). The use of higher-order factors representing overall measures of externalizing problem behaviors and social competence may provide a solution for dealing with the poor discriminant validity between some of the original SDQ subscales (Goodman et al., 2010; van Roy et al., 2008).

Contributions and Limitations

The current study adds important information on the SDQ's psychometric properties within an MTMM context, specifically focusing on this measure scores' reliability, as well as on their factorial, convergent, and discriminant validity. Our findings support the claim that researchers and practitioners should use multiple raters to achieve a more thorough evaluation of children's behavior using the SDQ. In particular, our results confirm the notion that parents and teachers provide partially unique information due to observation of different contexts (home vs. school). Teacher, mother, and father ratings of the SDQ seem to relate differently to children's age and sex, as well as to different nuances of children's self-regulation. For instance, we found that mother and father ratings mainly related to children's ability to comply with directions. However, teacher ratings were mostly associated with the ability to suppress a dominant response and initiate a subdominant response. We, therefore, recommend including at least one parent (mothers or fathers) and teachers in studies using the SDQ. In addition, we suggest that future studies be carried out to analyze the potential contributions of individual, relational, and contextual variables to convergent validity and method specificity of teacher, mother, and father ratings. Parents' demographic characteristics, the quality of the parent–child relationship and children's school year are some variables that may influence the shared and unique perspectives of teacher, mothers, and fathers regarding their children's psychological adjustment.

Although our findings are based on a robust analytical approach that allows estimating the convergent and discriminant validity coefficients while accounting for random measurement error, there are some limitations that should be noted. The participation rate was modest and the final sample was composed of children from dual-earner families, and included a large proportion of highly educated parents. These sample characteristics limit the generalization of the findings to children from other family configurations, clinical samples, and distinct age groups.

References

Achenbach, T., McConaughy, H., & Howel, C. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant

correlations for situational specificity. *Psychological Bulletin, 101,* 213–232. http://dx.doi.org/10.1037/0033-2909.101.2.213

Byrne, B., Shavelson, R., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105,* 456–466. http://dx.doi.org/10.1037/0033-2909.105.3.456

Caci, H., Morin, A., & Tran, A. (2015). Investigation of a bifactor model of the Strengths and Difficulties Questionnaire. *European Child & Adolescent Psychiatry, 24,* 1291–1301. http://dx.doi.org/10.1007/s00787-015-0679-3

Cadima, J., Enrico, M., Ferreira, T., Verschueren, K., Leal, T., & Matos, P. M. (2016). Self-regulation in early childhood: The interplay between family risk, temperament and teacher–child interactions. *European Journal of Developmental Psychology, 13,* 341–360. http://dx.doi.org/10.1080/17405629.2016.1161506

Campbell, D., & Fiske, D. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81–105. http://dx.doi.org/10.1037/h0046016

Cheung, G., & Rensvold, R. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9,* 233–255. http://dx.doi.org/10.1207/S15328007SEM0902_5

Coffman, D., & MacCallum, R. (2005). Using parcels to convert path analysis models into latent variable models. *Multivariate Behavioral Research, 40,* 235–259. http://dx.doi.org/10.1207/s15327906mbr4002_4

Davé, S., Nazareth, I., Senior, R., & Sherr, L. (2008). A comparison of father and mother report of child behaviour on the Strengths and Difficulties Questionnaire. *Child Psychiatry and Human Development, 39,* 399–413. http://dx.doi.org/10.1007/s10578-008-0097-6

De Los Reyes, A., Thomas, S., Goodman, K., & Kundey, S. (2013). Principles underlying the use of multiple informants' reports. *Annual Review of Clinical Psychology, 9,* 123–149. http://dx.doi.org/10.1146/annurev-clinpsy-050212-185617

Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika, 65,* 241–261. http://dx.doi.org/10.1007/BF02294377

Eid, M., Lischetzke, T., Nussbeck, F. W., & Trierweiler, L. I. (2003). Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple-indicator CT-C(M - 1) model. *Psychological Methods, 8,* 38–60. http://dx.doi.org/10.1037/1082-989X.8.1.38

Eid, M., Nussbeck, F., Geiser, C., Cole, D., Gollwitzer, M., & Lischetzke, T. (2008). Structural equation modeling of multitrait-multimethod data: Different models for different types of methods. *Psychological Methods, 13,* 230–253. http://dx.doi.org/10.1037/a0013219

Fan, Y., & Lance, C. (2017). A reformulated correlated trait–correlated method model for multitrait–multimethod data effectively increases convergence and admissibility rates. *Educational and Psychological Measurement, 77,* 1048–1063. http://dx.doi.org/10.1177/0013164416677144

Geiser, C., Burns, L., & Servera, M. (2014). Testing for measurement invariance and latent mean differences across methods: Interesting incremental information from multitrait-multimethod studies. *Frontiers in Psychology, 5,* 1–19. http://dx.doi.org/10.3389/fpsyg.2014.01216

Geiser, C., Eid, M., West, S., Lischetzke, T., & Nussbeck, F. (2012). A comparison of method effects in two confirmatory factor models for structurally different methods. *Structural Equation Modeling, 19,* 409–436. http://dx.doi.org/10.1080/10705511.2012.687658

Geiser, C., Mandelman, S., Tan, M., & Grigorenko, E. (2016). Multitrait–multimethod assessment of giftedness: An application of the correlated traits–correlated (methods − 1) model. *Structural Equation Modeling, 23,* 76–90. http://dx.doi.org/10.1080/10705511.2014.937792

Gomez, R. (2014). Correlated trait-correlated method minus one analysis of the convergent and discriminant validities of the Strengths and Difficulties Questionnaire. *Assessment, 21,* 372–382. http://dx.doi.org/10.1177/1073191112457588

Goodman, A., Lamping, D., & Ploubidis, G. (2010). When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): Data from British parents, teachers and children. *Journal of Abnormal Child Psychology, 38,* 1179–1191. http://dx.doi.org/10.1007/s10802-010-9434-x

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38,* 581–586. http://dx.doi.org/10.1111/j.1469-7610.1997.tb01545.x

Hill, C., & Hughes, J. (2007). An examination of the convergent and discriminant validity of the strengths and difficulties questionnaire. *School Psychology Quarterly, 22,* 380–406. http://dx.doi.org/10.1037/1045-3830.22.3.380

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55. http://dx.doi.org/10.1080/10705519909540118

Jöreskog, K. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34,* 183–202. http://dx.doi.org/10.1007/BF02289343

Jöreskog, K. (1971). Statistical analysis of sets of congeneric test. *Psychometrika, 36,* 109–133. http://dx.doi.org/10.1007/BF02291393

Kenny, D. (1976). An empirical application of confirmatory factor analysis to the multitrait-multimethod matrix. *Journal of Experimental Social Psychology, 12,* 247–252. http://dx.doi.org/10.1016/0022-1031(76)90055-X

Kersten, P., Czuba, K., McPherson, K., Dudley, M., Elder, H., Tauroa, R., & Vandal, A. (2016). A systematic review of evidence for the psychometric properties of the Strengths and Difficulties Questionnaire. *Methods and Measures, 40,* 64–75. http://dx.doi.org/10.1177/0165025415570647

Klein, A., Otto, Y., Fuchs, S., Zenger, M., & Von Klitzing, K. (2013). Psychometric properties of the parent-rated SDQ in preschoolers. *European Journal of Psychological Assessment, 29,* 96–104. http://dx.doi.org/10.1027/1015-5759/a000129

Kóbor, A., Takács, Á., & Urbán, R. (2013). The bifactor model of the strengths and difficulties questionnaire. *European Journal of Psychological Assessment, 29,* 299–307. http://dx.doi.org/10.1027/1015-5759/a000160

Kremer, P., Silva, A., De Cleary, J., Santoro, G., Weston, K., & Steele, E. (2015). Normative data for the Strengths and Difficulties Questionnaire for young children in Australia. *Journal of Paediatrics and Child Health, 51,* 970–975. http://dx.doi.org/10.1111/jpc.12897

Lance, C., Noble, C., & Scullen, S. (2002). A critique of the correlated trait—correlated method and correlated uniqueness models for multitrait—multimethod data. *Psychological Methods, 7,* 228–244. http://dx.doi.org/10.1037/1082-989X.7.2.228

Marzocchi, G., Capron, C., Di Pietro, M., Duran Tauleria, E., Duyme, M., Frigerio, A., . . . Thérond, C. (2004). The use of the Strengths and Difficulties Questionnaire (SDQ) in Southern European countries. *European Child & Adolescent Psychiatry, 13,* 40–46. http://dx.doi.org/10.1007/s00787-004-2007-1

Mellor, D., Wong, J., & Xu, X. (2011). Interparent agreement on the Strengths and Difficulties Questionnaire: A Chinese study. *Journal of Clinical Child and Adolescent Psychology, 40,* 890–896. http://dx.doi.org/10.1080/15374416.2011.614580

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741–749. http://dx.doi.org/10.1037/0003-066X.50.9.741

Mieloo, C., Raat, H., Van Oort, F., Bevaart, F., Vogel, I., Donker, M., & Jansen, W. (2012). Validity and reliability of the strengths and difficulties questionnaire in 5–6 year olds: Differences by gender or by parental

education? *PLoS ONE, 7,* e36805. http://dx.doi.org/10.1371/journal.pone.0036805

Muthén, B., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology, 25,* 267–316. http://dx.doi.org/10.2307/271070

Muthén, L., & Muthén, B. (2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Author.

Niclasen, J., Skovgaard, A., Andersen, A., Sømhovd, M., & Obel, C. (2013). A confirmatory approach to examining the factor structure of the Strengths and Difficulties Questionnaire (SDQ): A large scale cohort study. *Journal of Abnormal Child Psychology, 41,* 355–365. http://dx.doi.org/10.1007/s10802-012-9683-y

Oberski, D. (2014). lavaan.survey: An R Package for complex survey analysis of structural equation models. *Journal of Statistical Software, 57,* 1–27. http://dx.doi.org/10.18637/jss.v057.i01

Palmieri, P., & Smith, G. (2007). Examining the structural validity of the Strengths and Difficulties Questionnaire (SDQ) in a U.S. sample of custodial grandmothers. *Psychological Assessment, 19,* 189–198. http://dx.doi.org/10.1037/1040-3590.19.2.189

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rogge, J., Koglin, U., & Petermann, F. (2018). Do they rate in the same way? Testing of measurement invariance across parent and teacher SDQ ratings. *European Journal of Psychological Assessment, 34,* 69–78. http://dx.doi.org/10.1027/1015-5759/a000445

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48,* 1–36. http://dx.doi.org/10.18637/jss.v048.i02

Sanne, B., Torsheim, T., Heiervang, E., & Stormark, K. (2009). The strengths and difficulties questionnaire in the Bergen child study: A conceptually and methodically motivated structural analysis. *Psychological Assessment, 21,* 352–364. http://dx.doi.org/10.1037/a0016317

Satorra, A., & Bentler, P. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66,* 507–514. http://dx.doi.org/10.1007/BF02296192

Smith-Donald, R., Raver, C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the Preschool Self-regulation Assessment (PSRA) for field-based research. *Early Childhood Research Quarterly, 22,* 173–187. http://dx.doi.org/10.1016/j.ecresq.2007.01.002

Smits, I., Theunissen, M., Reijneveld, S., Nauta, M., & Timmerman, M. (2018). Measurement invariance of the parent version of the Strengths and Difficulties Questionnaire (SDQ) across community and clinical populations. *European Journal of Psychological Assessment, 34,* 238–246. http://dx.doi.org/10.1027/1015-5759/a000339

Stolk, Y., Kaplan, I., & Szwarc, J. (2017). Review of the strengths and difficulties questionnaire translated into languages spoken by children and adolescents of refugee background. *International Journal of Methods in Psychiatric Research, 26,* 1–21. http://dx.doi.org/10.1002/mpr.1568

Stone, L., Otten, R., Engels, R., Vermulst, A., & Janssens, J. (2010). Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-year-olds: A review. *Clinical Child and Family Psychology Review, 13,* 254–274. http://dx.doi.org/10.1007/s10567-010-0071-2

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software, 45,* 1–67. http://dx.doi.org/10.18637/jss.v045.i03

Van Leeuwen, K., Meerschaert, T., Bosmans, G., de Medts, L., & Braet, C. (2006). The strengths and difficulties questionnaire in a community sample of young children in Flanders. *European Journal of Psychological Assessment, 22,* 189–197. http://dx.doi.org/10.1027/1015-5759.22.3.189

Van Roy, B., Veenstra, M., & Clench-Aas, J. (2008). Construct validity of the five-factor Strengths and Difficulties Questionnaire (SDQ) in pre-, early, and late adolescence. *Journal of Child Psychology and Psychiatry, 49,* 1304–1312. http://dx.doi.org/10.1111/j.1469-7610.2008.01942.x

Vries, P., Davids, E., Mathews, C., & Aarø, L. (2018). Measuring adolescent mental health around the globe: Psychometric properties of the self-report strengths and difficulties questionnaire in South Africa, and comparison with U. K., Australian and Chinese data. *Epidemiology and Psychiatric Sciences, 27,* 369–380. http://dx.doi.org/10.1017/S2045796016001207

Widaman, K. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement, 9,* 1–26. http://dx.doi.org/10.1177/014662168500900101

Widaman, K., & Reise, S. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/10222-009

Widaman, K., & Thompson, J. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological Methods, 8,* 16–37. http://dx.doi.org/10.1037/1082-989X.8.1.16

Williamson, A., Mcelduff, P., Dadds, M., Este, C. D., Redman, S., Raphael, B., . . . Eades, S. (2014). The construct validity of the strengths and difficulties questionnaire for Aboriginal children living in urban New South Wales, Australia. *Australian Psychologist, 49,* 163–170. http://dx.doi.org/10.1111/ap.12045

Yu, J., Sun, S., & Cheah, C. (2016). Multitrait–multimethod analysis of the Strengths and Difficulties Questionnaire in young Asian American children. *Assessment, 23,* 603–613. http://dx.doi.org/10.1177/1073191115586459

Received August 14, 2019
Revision received August 5, 2020
Accepted August 26, 2020 ■