Article

Multitrait–Multimethod Analysis of the Strengths and Difficulties Questionnaire in Young Asian American Children

Assessment
2016, Vol. 23(5) 603–613

DOI: 10.1177/1073191115586459
asm.sagepub.com
SAGE

Jing Yu[1], Shuyan Sun[1], and Charissa S. L. Cheah[1]

Abstract

The Strengths and Difficulties Questionnaire (SDQ) is a widely used psychopathology screening tool that measures children's emotional symptoms, peer problems, conduct problems, hyperactivity/inattention, and prosocial behavior. Previous psychometric studies of the SDQ focused primarily on older children in Western cultures and suffered from several methodological limitations. This study examined the reliability, factor structure, convergent, and discriminant validity of the SDQ by focusing on young Asian American children and using more rigorous methods. The five-factor structure of the SDQ was confirmed by confirmatory factor analysis. The ω coefficients indicated adequate reliability for all subscales except parent-rated peer problems and conduct problems. The correlated trait–correlated method minus one multitrait–multimethod model provided evidence for convergent validity and discriminant validity of all subscales except for conduct problems relative to hyperactivity/inattention. This study provided new evidence for the psychometric properties of the SDQ in young children and cultural suitability of the SDQ for Asian Americans.

Keywords

Asian American children, confirmatory factor analysis, convergent and discriminant validity, multitrait–multimethod, Strengths and Difficulties Questionnaire, reliability

Screening children at an early age for mental health problems and delivering timely interventions are of great importance to prevent these childhood problems from progressing into severe psychiatric disorders (Harrington, Rutter, & Fombonne, 1996), especially for Asian Americans who tend not to seek professional help when they suffer mental illnesses (Sue, Cheng, Saad, & Chu, 2012). Thus, reliable and culturally valid screening tools are needed to detect emotional and behavioral problems of Asian American children early on. The Strengths and Difficulties Questionnaire (SDQ) is one of the most utilized screening instruments because it measures both psychosocial problems and strengths in children and provides quick detection of early childhood psychopathology (Goodman, 1997). The SDQ includes 25 items and was proposed to comprise five subscales—emotional symptoms, peer problems, conduct problems, hyperactivity/inattention, and prosocial behavior. Both parent and teacher versions of the SDQ (i.e., SDQP and SDQT) are available for children 3 to 16 years of age. A self-report version is also available for older children and adolescents aged 11 to 16 years (Goodman, Meltzer, & Bailey, 1998).

The SDQ has been translated into over 60 languages and widely used in developmental and clinical research in many different countries (e.g., Marzocchi et al., 2004; Ruchkin, Koposov, Vermeiren, & Schwab-Stone, 2012). Given that expectations and evaluations of child behaviors are culturally sensitive, it is crucial to empirically test, rather than assume, the applicability of SDQ in different cultural groups. However, despite extensive research on the psychometric properties of the SDQ in Western samples, psychometric studies among Asian or Asian American children are rare, especially young children. To justify the necessity of using more rigorous methods to examine the reliability, factor structure, convergent, and discriminant validity of the SDQ, findings from previous studies are reviewed below.

Reliability

According to a recent meta-analytic review of 48 studies (Stone, Otten, Engels, Vermulst, & Janssens, 2010), the internal reliability (measured by Cronbach's α) of the SDQT subscales was generally good (>.70; Cohen, 1977; Nunnally & Bernstein, 1994) except teacher-rated peer problems

[1]University of Maryland, Baltimore County, MD, USA

Corresponding Author:
Jing Yu, Department of Psychology, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA.
Email: jy6@umbc.edu

(.63), whereas the reliability of the SDQP subscales was generally low (.53 to .67) except parent-rated hyperactivity/inattention (.76). However, for any given subscale, there was a wide range of reliability estimates across studies. For instance, Du, Kou, and Coghill (2008) found that the reliability of four parent-rated subscales (.30 to .68) and three teacher-rated subscales (.48 to .63) were low in mainland Chinese children. This suggests that the SDQ may not be a reliable screening tool in some cultural contexts. Unfortunately, the meta-analysis by Stone et al. (2010) did not fully explore whether reliability varies across cultures due to the small number of studies conducted in non-Western samples.

Moreover, Cronbach's α has been heavily criticized as the solitary indicator of reliability because it usually underestimates true reliability (Bentler, 2009; Sijtsma, 2009) and may not be an indicator of scale unidimensionality (Schmitt, 1996). Among several alternative measures to α, McDonald's ω is the preferred estimate of internal reliability in many situations (Zinbarg, Revelle, Yovel, & Li, 2005). Defined as the proportion of variance in the scale score that is due to a general factor, ω pertains to how well a scale measures one construct and is more sensitive to the internal structure of a measure than α (Revelle & Zinbarg, 2009). Only a few studies have examined the reliability of the SDQ using ω. Stone et al. (2013) reported ω of the SDQP among 9- to 12-year-old Dutch children and found adequate reliability for all parent-rated subscales (ranging from .74 to .91). Ezpeleta, Granero, de la Osa, Penelo, and Domènech (2013) reported ω for both SDQP (ranged from .67 to .82) and SDQT subscales (ranged from .75 to .93) in Spanish 3-year-old children, and the only reliability that was lower than .70 was for the parent-rated prosocial behavior subscale (.67). Thus, the extremely low reliability of the SDQ among Chinese children (Du et al., 2008) may be attributed to the cultural unsuitability of the measure and/or the underestimate of reliability by α. The reliability of the SDQ using ω needs further evaluation in children of different cultures and age periods, including Asian American preschool children.

Factor Structure

The factor structure of the SDQ has been widely studied, but the statistical methods used in prior study were often not adequate. Many studies used principal component analysis or exploratory factor analysis (e.g., Du et al., 2008; Muris, Meesters, & van den Berg, 2003) to explore the factor structure underlying the SDQ. However, Goodman (1997) proposed the five dimensions of the SDQ based on theoretical foundations of child psychopathology. Therefore, confirmatory factor analysis (CFA) is more suitable than exploratory analyses. Moreover, findings from previous studies are not consistent. The CFA evidence for the five-factor structure was generally supported (see a review of previous studies by Stone et al., 2010; see also A. Goodman, Lamping, & Ploubidis, 2010; Klein, Otto, Fuchs, Zenger, & von Klitzing, 2013; McCrory & Layte, 2012; Mieloo et al., 2012), but some studies (e.g., Dickey & Blumberg, 2004) suggested the substantive superiority of a three-factor structure with broader dimensions including internalizing, externalizing, and prosocial behaviors. Therefore, the five-factor structure of the SDQ should be further validated prior to investigating other aspects of its psychometric properties.

Convergent and Discriminant Validity

In addition to factor structure, convergent validity and discriminant validity are also important aspects of validity evaluation of a measure. Convergent validity refers to the extent to which the same trait is measured by different methods, whereas discriminant validity refers to the extent to which different traits in a measure are distinct (Van Roy, Veenstra, & Clench-Aas, 2008). Convergent and discriminant validity can be more robustly evaluated by using the multitrait–multimethod (MTMM) method. For the SDQ, traits are the different subscales on the SDQT and SDQP, and methods are the reporters, that is, teachers and parents. Using different reporters for the same construct to represent different measurement methods is common practice in MTMM analysis (Widaman, 2010). Different approaches to MTMM analysis and findings from previous studies for the SDQ are reviewed next.

The original MTMM approach (Campbell & Fiske, 1959) evaluates the correlation matrix of multiple traits (e.g., the five SDQ subscales) by multiple methods (e.g., parent- and teacher-report) using observed scores. Convergent validity is inferred when there are statistically significant and relatively large monotrait–heteromethod correlations (i.e., high correlations among the same traits across different methods). Discriminant validity is inferred when there are relatively low heterotrait–monomethod correlations (among different subscales within a method) or heterotrait–heteromethod correlations (among different traits across different methods). Two studies using the original MTMM approach found moderate to large convergent validity of the SDQ: Convergent validity coefficients (monotrait–heteromethod) ranged from .24 to .47 in A. Goodman et al. (2010) and from .26 to .47 in Hill and Hughes (2007). However, they provided mixed findings for the discriminant validity of the SDQ. Good discriminant validity was supported for most of the subscales in A. Goodman et al.'s (2010) study despite a few exceptions (e.g., poor discriminant validity between parent-rated hyperactivity/inattention and teacher-rated conduct problems), whereas the discriminant validity of the SDQ subscales in Hill and Hughes (2007) was generally poor.

Because the original MTMM approach uses observed scores that comprise measurement errors, the findings for the convergent and discriminant validity of the SDQ may be distorted (Green, Goldman, & Salovey, 1993). MTMM analysis conducted within a CFA framework has thus been developed and widely implemented (Eid et al., 2008). Two commonly applied models are the correlated trait–correlated method (CT-CM; Marsh, 1989) and the correlated trait–correlated uniqueness (CT-CU; Kenny, 1976) models. For both models, convergent validity is inferred if there are relatively high trait variances in the subscales and discriminant validity is inferred if there are low correlations between the latent trait factors. Method effect is inferred if there are relatively high variances for the different subscales within methods (for CT-CM model) or high correlations for the different error variances within methods (for CT-CU model).

Van Roy et al. (2008) used the CT-CM model to examine parent- and self-rated SDQ, and found evidence for convergent validity of SDQ subscales (support being higher for parent-rated hyperactivity/inattention). Hill and Hughes (2007) tried the CT-CM model for the parent and teacher versions in first-grade children but the model did not converge. They thus examined a CT-CU model and found moderate convergent validity and relatively poor discriminant validity of the SDQ. Moreover, both studies found considerable method effects, which weakens but does not eliminate support for the convergent validity of the SDQ in these studies. Despite the advantages of CT-CM and CT-CU models over the original MTMM approach, they suffer from theoretical or psychometric problems of their own (see a critique by Lance, Noble, & Scullen, 2002). The CT-CU model may provide biased estimates of trait factor loadings and correlations (Conway, Lievens, Scullen, & Lance, 2004), thus it is recommended to be invoked only when the CT-CM model fails. In fact, the CT-CM application often leads to inadmissible solutions (e.g., Hill & Hughes, 2007) and may confound a general trait factor with a method factor (Eid, Lischetzke, Nussbeck, & Trierweiler, 2003).

Derived from the CT-CM model, the correlated trait–correlated method minus one [CT-C($M-1$)] model (Eid, 2000) has been recommended as the preferred model for structurally different methods (Eid et al., 2008; Nussbeck, Eid, Geiser, Courvoisier, & Lischetzke, 2009). The basic philosophy underlying the CT-C($M-1$) model is that a trait cannot be measured independently of the method (Geiser, Eid, & Nussbeck, 2008). Thus, one of the methods is selected as the reference method, and the true scores of the reference method indicators, that is, the latent traits being estimated that are confounded with the reference method, are then used to predict the true scores of the nonreference methods indicators (Eid et al., 2003). The consistency coefficient represents the proportion of the true variance in the subscale that is explained by the reference method, with larger values indicating higher convergence of the nonreference method with the reference method. In contrast, the method-specificity coefficient indicates the proportion of the true variance in a subscale that is not shared with the reference method, with larger values indicating less convergent validity. Discriminant validity between the traits as reflected by the reference method is inferred when there are low latent trait correlations. Gomez (2014), so far the only application of CT-C($M-1$) on the SDQ, chose SDQP as the reference method and provided evidence for convergent validity between parent and teacher ratings on adolescents as well as discriminant validity of all subscales, except for the discriminant validity of conduct problems relative to hyperactivity/inattention. No studies have used the CFA–MTMM models to examine the validity of the SDQ in non-Western samples or young children.

The Present Study

Given the predominant focus on older children and adolescents in Western cultures and methodological limitations in previous studies, it is essential to further investigate the psychometric properties of the SDQ by focusing on young, non-Western children and using more rigorous statistical methods. Therefore, the purpose of the present study was to evaluate the reliability and validity of the SDQ in an Asian American preschool sample. Specifically, three psychometric issues were addressed: (a) confirming the five-factor structure of both SDQP and SDQT; (b) evaluating internal reliability by calculating ω coefficients for the SDQP and SDQT subscales, in addition to estimating Cronbach's α; (c) evaluating the convergent and discriminant validity of the SDQ using the CT-C($M-1$) CFA–MTMM model.

Method

Sample and Measure

Chinese and Korean immigrant parents ($N = 327$, $M_{age} = 36.91$ years, $SD = 4.37$) with preschool children ($M_{age} = 4.38$ years, $SD = 0.98$, 49.3% females) were recruited from various organizations (e.g., language schools, churches, day care centers, and grocery stores) in Maryland and Washington, D.C. area, representing two of the largest East Asian immigrant populations in the United States (U.S. Census Bureau, 2010). With the permission and assistance of the directors in these organizations, announcements were made to the parents regarding the study. Ethical approval for the study was obtained from the university institutional review board. The parents (97.8% mothers and 2.2% fathers) and school, church, or day care teachers rated the children's emotional and behavioral functioning on the SDQ (Goodman, 1997). Both SDQP (47.5% in Chinese,

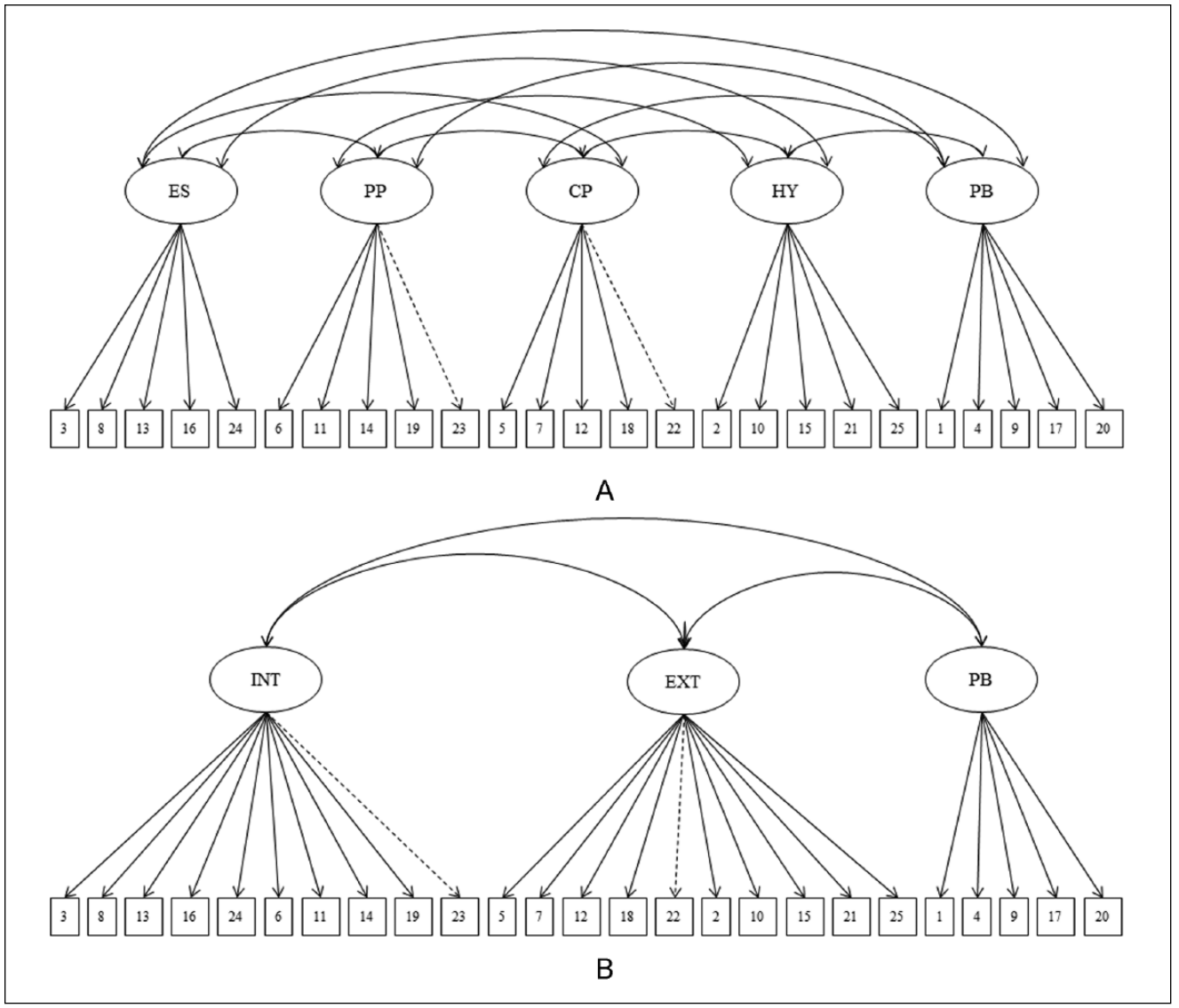

Figure 1. (A) The five-factor SDQ model. (B) The three-factor SDQ model.
Note. SDQ = Strengths and Difficulties Questionnaire; ES = emotional symptoms; PP = peer problems; CP = conduct problems; HY = hyperactivity/inattention; PB = prosocial behavior; INT = internalizing problem; EXT = externalizing problems. Item numbers are shown.

41.6% in Korean, and 10.9% in English) and SDQT (2.4% in Chinese, 27.8% in Korean, and 69.7% in English) include the same 25 items. Parents filled out the questionnaire during a home visit and teacher measures were obtained primarily by calling, faxing, or e-mailing. Each item was rated by the informant on a 3-point Likert-type scale (0 = *not true*, 1 = *somewhat true*, and 2 = *certainly true*). More than 96% of the children were from two-parent families. The immigrant parents have been in the United States for an average of 11 years (*SD* = 7.51), ranging from 2 months to 35 years. The majority (82.5%) of the parents had a college education or higher.

Analysis Plan

CFA was conducted to confirm the five-factor structure, in comparison with a three-factor structure, of the SDQ in the Asian American preschool sample (see Figure 1A and 1B for model diagrams). To ensure the cultural, linguistic, and/or informant equivalence of different versions of the SDQ, three sets of measurement invariance tests were conducted for SDQP and SDQT across the cultural and linguistic versions, as well as SDQ across the two informants (i.e., parents and teachers) using a single-sample approach (Ezpeleta & Penelo, 2015; Ferrando, 2000). After the factor analyses,

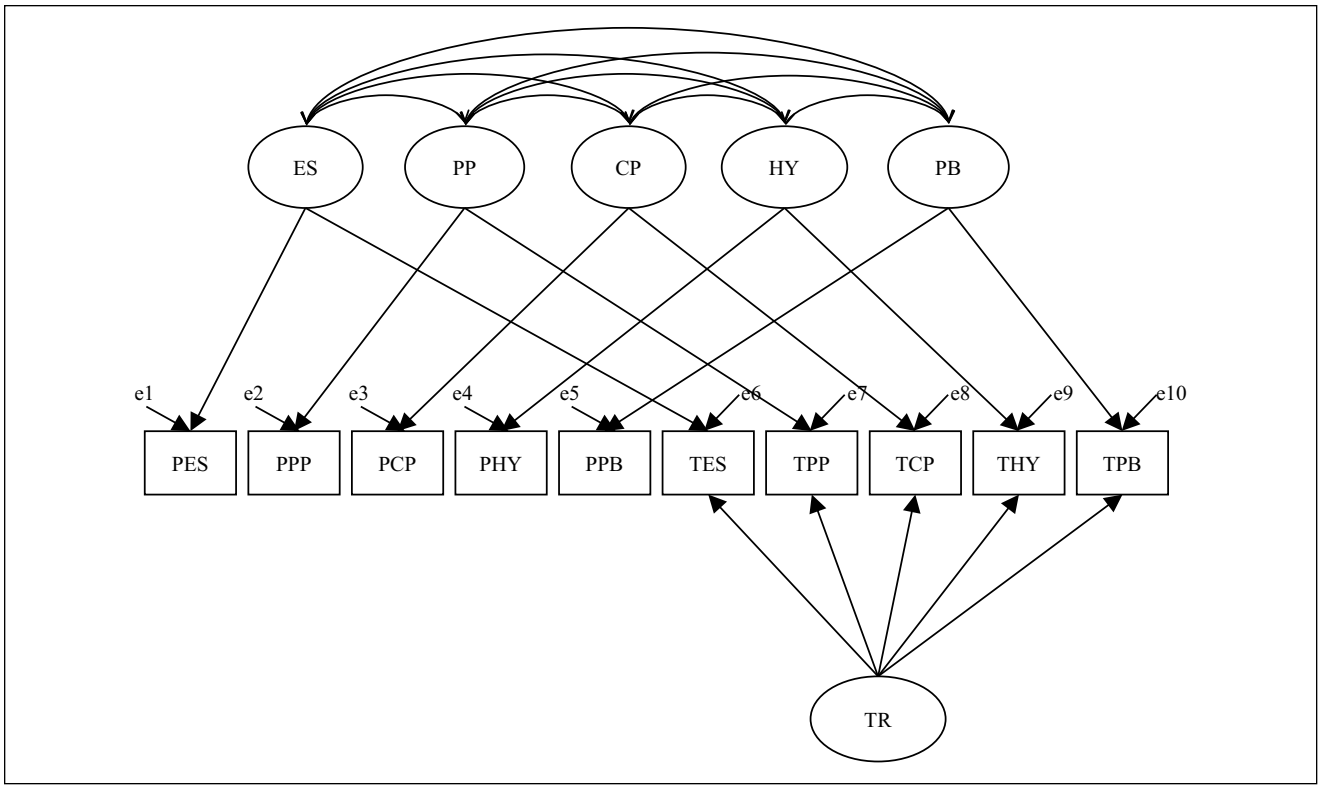

Figure 2. Hypothesized correlated trait–correlated method minus one [CT-C(*M* − 1)] model.
Note. P = parent-rated; T = teacher-rated; ES = emotional symptoms; PP = peer problems; CP = conduct problems; HY = hyperactivity/inattention; PB = prosocial behavior; TR = teacher report.

internal reliability was estimated based on the finally retained factor models of the SDQ. Cronbach's α were obtained from SPSS, whereas ω coefficients were calculated according to the formula provided by Revelle and Zinbarg (2009) and Stone et al. (2013). A CT-C(*M* − 1) model (see Figure 2) was then used to evaluate the convergent and discriminant validity of the SDQ. To facilitate cross-study comparison with Gomez (2014), the SDQP was also chosen as the reference method in this study. In line with the model specifications, all indicators of the reference method (i.e., all subscales of the SDQP) loaded on their appropriate underlying latent traits but not on any method factor. The indictors of the nonreference method were linked both to the appropriate trait factors and to a method factor. The trait factors correlate with each other but do not correlate with the method factor.

The CFA models were analyzed in M*plus* 7 (Muthén & Muthén, 1998-2012). To confirm the factor structure of SDQ, the weighted least squares estimator with mean and variance adjusted was used because the item data were ordinal in nature. For the MTMM model, robust maximum likelihood estimator was used to deal with the violation of multivariate normality in continuous data. The Satorra–Bentler robust scaled chi-square statistic (S-Bχ^2), the comparative fit index (CFI), the root mean squared error of approximation (RMSEA), and the standardized root mean square residual (SRMR) were used to evaluate the overall model fit (SRMR was not applicable to the CFA models because weighted least squares estimator with mean and variance adjusted was used). Good model fit is evidenced by CFI above .95, RMSEA below .06, and SRMR below .08, whereas acceptable model fit requires CFI above .90, RMSEA below .08, and SRMR below .10 (Brown, 2006; Hu & Bentler, 1998). The full-information maximum likelihood estimation (Enders, 2013) was used to handle missing data (less than 0.2% of the data were missing in this study). The magnitude of all correlations was interpreted based on the guideline proposed by Hemphill (2003): small (<.20), moderate (.20-.30), and large (>.30). A latent correlation <.85 was considered supportive of discriminant validity between the traits in CFA–MTMM framework (Brown, 2006; Gomez, 2014).

Results

Factor Structure and Measurement Invariance

The original five-factor model and a reduced dimensionality of three-factor model were estimated for both SDQP and SDQT. For the original five-factor SDQP model, the RMSEA

(.05) indicated good model fit, but the CFI (.87) was slightly lower than the acceptable fit. All factor loadings were significant except Item 23 (λ = −.01). Item 19 (λ = .26) and Item 22 (λ = .28) had standardized factor loadings lower than .40, but the rest of the items had absolute loadings ranging from .42 to .82. The three-factor SDQP model (RMSEA = .06 and CFI = .82) fit the data significantly worse than the five-factor model, $\Delta\chi^2(7, N = 327) = 74.39$, $p < .001$. The original five-factor SDQT model could not converge with Item 22 in the model. After removing Item 22, the SDQT model achieved acceptable model fit, RMSEA = .07and CFI = .90. All factor loadings were statistically significant and had absolute loadings ranging from .43 to .96, with Item 23 (λ = .26) being the only exception. Similar to the SDQP, the three-factor SDQT model (RMSEA = .08 and CFI = .86) fit the data significantly worse than the three-factor model, $\Delta\chi^2(7, N = 327) = 106.73$, $p < .001$. Therefore, five-factor SDQ models were preferred over the three-factor models.

Due to the model nonconvergence issue (caused by Item 22 as per the warning message produced by M*plus*) and the low factor loading of Item 23, both items were removed from the five-factor models. This ensured that the same items in SDQP and SDQT be used for subsequent analysis. The trimmed five-factor models (i.e., without Items 22 and 23) did not achieve acceptable model fit (SDQP: RMSEA = .06 and CFI = .88; SDQT: RMSEA = .07 and CFI = .91). We then allowed correlations between the unique variances of some individual items based on modification indices. This would not change the substantive conclusions regarding the adequacy of the hypothesized factor structure (Bollen, 1989; A. Goodman et al., 2010). The final five-factor models achieved good or acceptable model fit (SDQP: RMSEA = .05 and CFI= .91; SDQT: RMSEA = .06 and CFI = .92).

Tests of measurement invariance were conducted to ensure equivalence across cultural, linguistic, or informant versions of the SDQ. Metric invariance (i.e., invariant factor loadings) was established for (a) SDQP across Chinese and Korean groups, $\Delta\chi^2(23, N = 327) = 34.08$, $p = .06$; (b) SDQT across Chinese and Korean groups, $\Delta\chi^2(23, N = 327) = 32.96$, $p = .08$; and (c) SDQ across parent and teacher versions, $\Delta\chi^2(18, N = 327) = 22.97$, $p = .19$. Standardized factor loadings for the original and final five-factor SDQP and SDQT models are presented in Table 1.

Reliability

Descriptive statistics, correlations, and reliability estimates for the SDQ subscales were presented in Table 2. Cronbach's α and ω coefficients were estimated based on the final factor models of the SDQP and SDQT. According to Cronbach's α, parent-rated hyperactivity/inattention (.73) and prosocial behavior (.70) had acceptable reliability and the other three subscales had low reliability (<.60). For SDQT, the reliability of peer problems and conduct problems was also low (<.60), but acceptable reliability was achieved for other dimensions: emotional symptoms (.70), hyperactivity/inattention (.80), and prosocial behavior (.79). According to ω coefficients, only parent-rated peer problems and conduct problems had low reliability (<.70), whereas the rest of three parent-rated subscales and all five teacher-rated subscales achieved acceptable to good reliability (ranging from .72 to .88).

Convergent and Discriminant Validity

Based on the final SDQ factor models, sum scores of the five subscales in each version were computed and used in the subsequent MTMM analysis. Based on the traditional MTMM correlation matrix approach (Campbell & Fiske, 1959) shown in Table 2, monotraits–heteromethod correlations ranged from .18 (conduct problems) to .37 (peer problems), indicating low to moderate convergent validity. The large heterotrait–monomethod correlations suggested that prosocial behavior did not discriminate well from peer problems, conduct problems, and hyperactivity/inattention (*r*s ranging from −.35 to −.47); neither did conduct problems relative to hyperactivity/inattention ($r = .40$ for SDQP and $r = .53$ for SDQT).

The convergent and discriminant validity were further evaluated using the CFA–MTMM analysis. The fit indices for the CT-C($M - 1$) model were S-B$\chi^2(20, N = 327) = 64.16$, $p < .001$, CFI = .93, RMSEA = .08, and SRMR = .04. The CFI and RMSEA suggested acceptable model fit and the SRMR. indicated good fit. By using formulas in Eid et al. (2003), consistency and method-specificity coefficients for both observed and true-score variables were calculated. The latent correlation between the true scores of the reference and nonreference methods for the same trait was calculated by taking the square root of the consistency coefficient of the true-score variable. Convergent validity is inferred if a subscale has a statistically significant consistency coefficient for the observed variables and large latent correlation between true-score variables (Gomez, 2014). Preferably, the subscale has a larger consistency than method-specificity coefficient to indicate stronger support for convergent validity. Discriminant validity between the traits is inferred if there are low correlations between different latent traits estimated by the reference method. The variance components and latent correlations between parent- and teacher-rating on the five traits were presented in Table 3.

In the CT-C($M - 1$) model, all consistency coefficients for the observed variables were statistically significant, and there were large latent correlations for all the traits (latent correlations ranging from .40 to .75), supporting the convergent validity of mother–teacher ratings of the SDQ subscales. Additionally, emotional symptoms and peer problems had larger consistency coefficients than method-specificity coefficients, whereas conduct problems, hyperactivity/inattention, and prosocial behavior had larger

Table 1. Standardized Factor Loadings for SDQ Items in Original and Finally Retained Factor Models.

		Original model		Final model	
Subscales	Items	SDQP	SDQT	SDQP	SDQT
Emotional symptoms	3. Somatic	.42	.72	.42	.72
	8. Worries	.61	.63	.60	.63
	13. Unhappy	.82	.72	.82	.72
	16. Clingy	.58	.76	.58	.76
	24. Fears	.55	.80	.56	.81
Peer problems	6. Solitary	.48	.43	.48	.40
	11r. Good friend	−.67	−.68	−.67	−.66
	14r. Popular	−.81	−.85	−.81	−.83
	19. Bullied	.26	.59	.26	.57
	23. Best with adults	−.01[a]	.26	N/A[b]	N/A[b]
Conduct problems	5. Tempers	.62	.64	.62	.64
	7r. Obedient	−.51	−.96	−.50	−.96
	12. Fights	.67	.65	.66	.65
	18. Lies	.42	.43	.39	.43
	22. Steals	.28	N/A[b]	N/A[b]	N/A[b]
Hyperactivity/inattention	2. Restless	.55	.83	.55	.78
	10. Fidgety	.72	.73	.72	.65
	15. Distractible	.71	.78	.71	.79
	21r. Reflective	−.78	−.71	−.79	−.71
	25r. Persistent	−.77	−.89	−.77	−.89
Prosocial behavior	1. Considerate	.67	.86	.67	.87
	4. Shares	.66	.74	.66	.75
	9. Caring	.73	.81	.73	.73
	17. Kind to kids	.51	.65	.50	.65
	20. Help out	.77	.77	.77	.68

Note. SDQ = Strengths and Difficulties Questionnaire; SDQP = Strengths and Difficulties Questionnaire–Parent version; SDQT = Strengths and Difficulties Questionnaire–Teacher version.
[a]The only item that had nonsignificant factor loading. [b]N/A because item was not included in the factor model.

method-specificity coefficients than consistency coefficients. These findings indicated more support for the convergence of mother–teacher ratings for emotional symptoms and peer problems compared with other subscales. As for discriminant validity, with the exception of the latent correlation between conduct problems and hyperactivity/inattention (.87), all other correlations between the latent traits were smaller than .85 despite their moderate to large magnitudes (absolute values ranging from .22 to .68). Thus, support was evidenced for the discriminant validity between all trait factors except between conduct problems and hyperactivity/inattention.

Discussion

Factor Structure

The results of the present study provided some support for the proposed five-factor structure of the SDQ in an Asian American preschool sample. However, several of the findings deserve closer examination. Two items were particularly problematic: Item 22 (Steals) caused model unconvergence in the SDQT model and Item 23 (Best with adults) had a nonsignificant factor loading in the SDQP model. Therefore, both were removed from the subsequent MTMM analysis. Item 22 was also removed in some Western studies because of its extreme skewness (Kóbor, Takács, & Urbán, 2013) or failure to load on any factors (Dickey & Blumberg, 2004), but worked well in most Western studies and a Chinese study (Liu et al., 2013). Thus, there is no clear pattern on whether Item 22 behaves well in different cultural contexts. As for Item 23, although it performed generally well in Western studies, two previous Chinese studies (Du et al., 2008; Liu et al., 2013) as well as this study did not find significant factor loadings. This suggests that Item 23 might be more culturally dependent. It is possible that "gets along better with adults than with other children" is not necessarily an abnormal phenomenon for young children in the eyes of Chinese or Korean parents due to the greater focus on parent-child emotional interdependence in

Table 2. Correlations, Descriptive Statistics, and Internal Reliability Estimates of the SDQ Scales.

	PES	PPP	PCP	PHY	PPB	TES	TPP	TCP	THY	TPB
PES	—	.31**	.23**	.14*	−.17**	**.23****	.09	−.01	−.10	−.02
PPP		—	.29**	.16**	−.45**	.15**	**.37****	.16**	.17**	−.28**
PCP			—	.44**	−.35**	.04	.20**	**.18****	.13*	−.12*
PHY				—	−.36**	.09	.13*	.20**	**.23****	−.20**
PPB					—	−.09	−.26**	−.21**	−.21**	**.34****
TES						—	.35**	.22**	.21**	−.14*
TPP							—	.26**	.39**	−.47**
TCP								—	.53**	−.36**
THY									—	−.45**
TPB										—
M	1.84	1.40	1.63	3.50	6.86	1.29	1.13	0.82	2.72	7.11
SD	1.61	1.36	1.27	2.17	2.04	1.67	1.36	1.16	2.43	2.36
Skewness	0.95	0.97	1.03	0.56	−0.21	1.46	1.37	1.74	0.83	−0.55
Kurtosis	0.58	1.00	1.86	−0.08	−0.54	1.68	1.87	3.40	0.01	−0.46
Cronbach's α	.55	.50	.45	.73	.70	.70	.55	.56	.80	.79
ω coefficient	.74	.65	.63	.80	.80	.85	.72	.78	.88	.86

Note. SDQ = Strengths and Difficulties Questionnaire; P = parent-rated; T = teacher-rated; ES = emotional symptoms; PP = peer problems; CP = conduct problems; HY = hyperactivity/inattention; PB = prosocial behavior. Subscale scores were calculated based on the final factor models after removing Items 22 and 23 for multitrait–multimethod analysis. Correlations in bold indicate convergent validity coefficients (monotrait–heteromethod); Underlined correlations indicate discriminant validity coefficients (heterotrait–monomethod).
*$p < .05$. **$p < .01$.

Table 3. Variance Components in the Correlated Trait–Correlated Method Minus One Model With SDQP as the Reference Method.

	Observed variables			True-score variables		
	Reliability	Consistency	Method-specificity	Consistency	Method-specificity	Latent correlation
Emotional symptoms						
Mother	.51	.51**	.00	1.00	.00	1.00
Teacher	.27	.15**	.12	.56	.44	.75
Peer problems						
Mother	.53	.53**	.00	1.00	.00	1.00
Teacher	.43	.22**	.21	.51	.49	.71
Conduct problems						
Mother	.48	.48**	.00	1.00	.00	1.00
Teacher	.42	.10**	.32	.24	.76	.49
Hyperactivity/inattention						
Mother	.52	.52**	.00	1.00	.00	1.00
Teacher	.61	.10**	.51	.16	.84	.40
Prosocial behavior						
Mother	.55	.55**	.00	1.00	.00	1.00
Teacher	.44	.19**	.25	.44	.56	.67

Note. SDQ = Strengths and Difficulties Questionnaire; SDQP = Strengths and Difficulties Questionnaire–Parent version. Reliability coefficient is the sum of the consistency and method-specificity coefficient for a given observed variable, which is also an underestimation of reliability (Eid et al., 2003). Thus, ω coefficients, instead of reliability coefficients from the correlated trait–correlated method minus one model were primarily interpreted as indicators of internal reliability for the SDQ subscales. True-score variables referred to the latent traits measured by the reference method (i.e., SDQP in this study).
**$p < .01$.

Asian cultures (Chao & Tseng, 2002). This finding calls for an inspection of the translation and conceptual clarity of this particular item in the Chinese and Korean versions of the SDQP.

Another issue lies in the considerable overlap among several subscales of the SDQ. Large correlations between hyperactivity/inattention and conduct problems indicated that they may partly share the same underlying trait, similarly for

emotional symptoms and peer problems. These findings are consistent with the three-factor structure of SDQ found by Dickey and Blumberg (2004) and the five-factor second order model found by A. Goodman et al. (2010). The externalizing dimension is likely to be responsible for the conceptual overlap between hyperactivity/inattention and conduct problems, and the internalizing dimension may account for the overlap between emotional symptoms and peer problems. Furthermore, consistent with the findings in Goodman (2001), our study indicated that internalizing subscales were not substantially correlated with externalizing subscales except the correlation between teacher-rated hyperactivity/inattention and peer problems (r = .39). However, our model comparison results suggested that the five-factor structure was superior to the three-factor structure in terms of model fit. Therefore, together with A. Goodman et al. (2010), we advocate the use of five-factor first or second order models instead of the three-factor structure.

Reliability

The low reliability as estimated by α for parent-rated conduct problems, peer problems, and emotional symptoms as well as teacher-rated conduct problems and peer problems was largely consistent with previous studies (Goodman, 2001; Palmieri & Smith, 2007). However, based on ω coefficients, all SDQT subscales and three SDQP subscales had adequate to good internal reliability, indicating the reliability of SDQ might indeed be underestimated in Chinese samples by using α (e.g., Du et al., 2008). On the other hand, when comparing with the study of Ezpeleta et al. (2013) that reported ω for both SDQP and SDQT in a Spanish preschool sample, despite much comparability, we found lower ω for parent-rated peer problems (.65 vs. .70) and conduct problems (.63 vs. .72), and much higher ω for prosocial behavior (.80 vs. .67). Nevertheless, all SDQP subscales were reliable (ωs > .70) in Dutch children (Stone et al., 2013). These emerged differences in the three samples imply cultural variations in the reliability of parent-rated SDQ subscales. Thus, researchers cannot assume the reliability of SDQP in a specific culture without testing it using an appropriate measure (e.g., the ω coefficient).

Convergent and Discriminant Validity

Based on the original MTMM matrix, support for the convergent validity of mother–teacher ratings and partial support for the discriminant validity of SDQ subscales was found in our sample of Asian American preschool children, largely consistent with A. Goodman et al. (2010), which took the same approach. The moderate level of mother–teacher agreement also corroborated previous findings (e.g., Goodman, 2001; Sanne, Torsheim, Heiervang, & Stormark, 2009; Van Leeuwen, Meerschaert, Bosmans, De Medts, & Braet, 2006). Based on the CT-C(M − 1) model, support for both convergent and discriminant validity was found except that conduct problems did not discriminate well from hyperactivity/inattention (Gomez, 2014). This finding is different from Hill and Hughes (2007), which found poor discriminant validity using the CT-CU model. The mixed findings of discriminant validity are likely due to different criteria used to interpret the observed and latent correlations as evidence of discriminant validity when using different MTMM models. Furthermore, as noted by Van Roy et al. (2008), the high cross-scale correlations in younger children are not surprising because possible emotional and behavioral difficulties are less distinct early in life and many children have problems in different areas at the same time.

Method factors tend to make large contributions to psychological measures, including the SDQ. Van Roy et al. (2008) found that the method factor in their CT-CM model had little influence on the hyperactivity/inattention subscale but did affect other SDQP subscales. The present study also found method effect as evidenced by salient method-specificity coefficients in the CT-C(M − 1) model. Therefore, using multiple methods to measure the same construct may be better than relying on a single method. Our findings, along with Goodman (2001), underscore the importance of using multi-informants of the SDQ in the assessment of children's mental health so that the psychometric properties of the SDQ can be appropriately evaluated prior to using it to screen children's emotional and behavioral problems.

Limitations and Contributions

Several limitations of the present study need to be addressed. The study sample was composed of two East Asian immigrant mothers with high parental education levels. Although generally representative of the larger first-generation Asian immigrant population in the Maryland-D.C. metropolitan area and other emerging immigrant areas (McCabe, 2012), the findings may not be generalizable to other Chinese and Korean immigrant populations in the United States or other Asian immigrant families originating from different Asian countries. Moreover, the findings in the present study were based on a community sample of young children; thus, the utility of the SDQ in clinical samples of Asian American children remains unclear and needs to be investigated (Lai et al., 2010; Van Leeuwen et al., 2006).

Despite the limitations, this study used ω coefficients in addition to Cronbach's α, providing new insight for the reliability estimates of the SDQ, especially for a non-Western sample. Given that ω coefficient is a better estimate of internal reliability than α, future studies should either report both coefficients or use ω as an alternative to α. Moreover, this study is among the first to examine the convergent and discriminant validity of the SDQ in young Asian American children using a CFA–MTMM approach that has the advantage of reducing measurement error and estimating the convergent and discriminant correlations

between the hypothesized latent traits (Brown, 2006; A. Goodman et al., 2010). Together, these findings added important evidence for the psychometric properties of SDQ and provided a strong empirical basis for the continued use of SDQ in applied and research work with Asian American children.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by National Institute of Child Health, 1R03HD052827-01 awarded to Charissa S. L. Cheah.

References

Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika*, *74*, 137-143. doi:10.1007/s11336-008-9100-1

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

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

Chao, R., & Tseng, V. (2002). Parenting of Asians. In M. H. Bornstein (Ed.), *Handbook of parenting: Social conditions and applied parenting* (2nd ed., Vol. 4, pp. 59-93). Mahwah, NJ: Lawrence Erlbaum.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (Rev. ed.). New York, NY: Academic Press.

Conway, J. M., Lievens, F., Scullen, S. E., & Lance, C. E. (2004). Bias in the correlated uniqueness model for MTMM data. *Structural Equation Modeling*, *11*, 535-559. doi:10.1207/s15328007sem1104_3

Dickey, W. C., & Blumberg, S. J. (2004). Revisiting the factor structure of the strengths and difficulties questionnaire: United States, 2001. *Journal of the American Academy of Child & Adolescent Psychiatry*, *43*, 1159-1167. doi:10.1097/01.chi.0000132808.36708.a9

Du, Y., Kou, J., & Coghill, D. (2008). The validity, reliability and normative scores of the parent, teacher and self report versions of the Strengths and Difficulties Questionnaire in China. *Child and Adolescent Psychiatry and Mental Health*, *2*(1), 8-22. doi:10.1186/1753-2000-2-8

Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika*, *65*, 241-261. doi: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. doi:10.1037/1082-989X.8.1.38

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

Enders, C. K. (2013). Dealing with missing data in developmental research. *Child Development Perspectives*, *27*, 27-31. doi:10.1111/cdep.12008

Ezpeleta, L., Granero, R., de la Osa, N., Penelo, E., & Domènech, J. M. (2013). Psychometric properties of the Strengths and Difficulties Questionnaire in 3-year-old preschoolers. *Comprehensive Psychiatry*, *54*, 282-291. doi:10.1016/j.comppsych.2012.07.009

Ezpeleta, L., & Penelo, E. (2015). Measurement invariance of oppositional defiant disorder dimensions in 3-year-old preschoolers. *European Journal of Psychological Assessment*, *31*, 45-53. doi:10.1027/1015-5759/a000205

Ferrando, P. J. (2000). Testing the equivalence among different item response formats in personality measurement: A structural equation modeling approach. *Structural Equation Modeling*, *7*, 271-286. doi:10.1207/S15328007SEM0702_7

Geiser, C., Eid, M., & Nussbeck, F. W. (2008). On the meaning of the latent variables in the CT-C(*M*-1) model: A comment on Maydeu-Olivares and Coffman (2006). *Psychological Methods*, *13*, 49-57. doi:10.1037/1082-989X.13.1.49

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. doi:10.1177/1073191112457588

Goodman, A., Lamping, D. L., & Ploubidis, G. B. (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. doi: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. doi:10.1111/j.1469-7610.1997.tb01545.x

Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*, 1337-1345. doi:10.1097/00004583-200111000-00015

Goodman, R., Meltzer, H., & Bailey, V. (1998). The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. *European Child & Adolescent Psychiatry*, *7*, 125-130. doi:10.1007/s007870050057

Green, D. P., Goldman, S. L., & Salovey, P. (1993). Measurement error masks bipolarity in affect ratings. *Journal of Personality and Social Psychology*, *64*, 1029-1041. doi:10.1037/0022-3514.64.6.1029

Harrington, R., Rutter, M., & Fombonne, E. (1996). Developmental pathways in depression: Multiple meanings, antecedents, and endpoints. *Development and Psychopathology*, *8*, 601-616. doi:10.1017/S095457940000732X

Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist*, *58*, 78-79. doi:10.1037/0003-066X.58.1.78

Hill, C. R., & Hughes, J. N. (2007). An examination of the convergent and discriminant validity of the Strengths and Difficulties Questionnaire. *School Psychology Quarterly*, *22*, 380-406. doi:10.1037/1045-3830.22.3.380

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model

misspecification. *Psychological Methods*, *3*, 424-453. doi:10.1037/1082-989X.3.4.424

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

Klein, A. M., 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. doi: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. doi:10.1027/1015-5759/a000160

Lai, K. Y., Luk, E. S., Leung, P. W., Wong, A. S., Law, L., & Ho, K. (2010). Validation of the Chinese version of the strengths and difficulties questionnaire in Hong Kong. *Social Psychiatry and Psychiatric Epidemiology*, *45*, 1179-1186. doi:10.1007/s00127-009-0152-z

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

Liu, S. K., Chien, Y. L., Shang, C. Y., Lin, C. H., Liu, Y. C., & Gau, S. S. F. (2013). Psychometric properties of the Chinese version of Strength and Difficulties Questionnaire. *Comprehensive Psychiatry*, *54*, 720-730. doi:10.1016/j.comppsych.2013.01.002

Marsh, H. W. (1989). Confirmatory factor analyses of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement*, *13*, 335-361. doi:10.1177/014662168901300402

Marzocchi, G. M., Capron, C., Di Pietro, M., Tauleria, E. D., 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*(2), ii40-ii46. doi:10.1007/s00787-004-2007-1

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

McCabe, K. (2012). *Chinese immigrants in the United States. Migration information source* (ISSN No. 1946-40370). Washington, DC: Migration Policy Institute. Retrieved from http://www.migrationinformation.org/USfocus/display.cfm?id=876

McCrory, C., & Layte, R. (2012). Testing competing models of the Strengths and Difficulties Questionnaire's (SDQ's) factor structure for the parent-informant instrument. *Personality and Individual Differences*, *52*, 882-887. doi:10.1016/j.paid.2012.02.011

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. doi:10.1371/journal.pone.0036805

Muris, P., Meesters, C., & van den Berg, F. (2003). The Strengths and Difficulties Questionnaire (SDQ). *European Child & Adolescent Psychiatry*, *12*, 1-8. doi:10.1007/s00787-003-0298-2

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.

Nussbeck, F. W., Eid, M., Geiser, C., Courvoisier, D. S., & Lischetzke, T. (2009). A CTC(*M*-1) model for different types of raters. *Methodology*, *5*, 88-98. doi:10.1027/1614-2241.5.3.88

Palmieri, P. A., & Smith, G. C. (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. doi:10.1037/1040-3590.19.2.189

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, *74*, 145-154. doi:10.1007/s11336-008-9102-z

Ruchkin, V., Koposov, R., Vermeiren, R., & Schwab-Stone, M. (2012). The Strength and Difficulties Questionnaire: Russian validation of the teacher version and comparison of teacher and student reports. *Journal of Adolescence*, *35*, 87-96. doi:10.1016/j.adolescence.2011.06.003

Sanne, B., Torsheim, T., Heiervang, E., & Stormark, K. M. (2009). The Strengths and Difficulties Questionnaire in the Bergen Child Study: A conceptually and methodically motivated structural analysis. *Psychological Assessment*, *21*, 352-364. doi:10.1037/a0016317

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*, 350-353. doi:10.1037/1040-3590.8.4.350

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*, 107-120. doi:10.1007/s11336-008-9101-0

Stone, L. L., Otten, R., Engels, R. C., Vermulst, A. A., & Janssens, J. M. (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. doi:10.1007/s10567-010-0071-2

Stone, L. L., Otten, R., Ringlever, L., Hiemstra, M., Engels, R. C., Vermulst, A. A., & Janssens, J. M. (2013). The parent version of the Strengths and Difficulties Questionnaire: Omega as an alternative to alpha and a test for measurement invariance. *European Journal of Psychological Assessment*, *29*, 44-50. doi:10.1027/1015-5759/a000119

Sue, S., Cheng, J. K. Y., Saad, C. S., & Chu, J. P. (2012). Asian American mental health: A call to action. *American Psychologist*, *67*, 532-544. doi:10.1037/a0028900

U.S. Census Bureau. (2010). Overview of Race and Hispanic Origin: 2010. *2010 Census Briefs*. Retrieved from http://www.census.gov/prod/cen2010/briefs/c2010br-02.pdf.

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. doi: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. doi:10.1111/j.1469-7610.2008.01942.x

Widaman, K. F. (2010). Multitrait-multimethod analysis. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 299-314). New York, NY: Routledge.

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω_H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, *70*, 123-133. doi:10.1007/s11336-003-0974-7