Parent–Teacher Agreement on Children's Problems in 21 Societies

Leslie A. Rescorla , Lauren Bochicchio , Thomas M. Achenbach , Masha Y. Ivanova , Fredrik Almqvist , Ivan Begovac , Niels Bilenberg , Hector Bird , Anca Dobrean , Nese Erol , Eric Fombonne , Antonio Fonseca , Alessandra Frigerio , Daniel S. S. Fung , Michael C. Lambert , Patrick W. L. Leung , Xianchen Liu , Ivica Marković , Jasminka Markovic , Asghar Minaei , Yoon Phaik Ooi , Alexandra Roussos , Vlasta Rudan , Zeynep Simsek , Jan van der Ende , Sheila Weintraub , Tomasz Wolanczyk , Bernardine Woo , Bahr Weiss , John Weisz , Rita Zukauskiene & Frank C. Verhulst

R Routledge
Taylor & Francis Group

CROSS-CULTURAL WORK

Parent–Teacher Agreement on Children's Problems in 21 Societies

Leslie A. Rescorla[1], Lauren Bochicchio[2], Thomas M. Achenbach[3], Masha Y. Ivanova[3], Fredrik Almqvist[4], Ivan Begovac[5], Niels Bilenberg[6], Hector Bird[7], Anca Dobrean[8], Nese Erol[9], Eric Fombonne[10], Antonio Fonseca[11], Alessandra Frigerio[12], Daniel S. S. Fung[13], Michael C. Lambert[14], Patrick W. L. Leung[15], Xianchen Liu[16], Ivica Marković[17], Jasminka Markovic[18], Asghar Minaei[19], Yoon Phaik Ooi[13], Alexandra Roussos[20], Vlasta Rudan[5], Zeynep Simsek[21], Jan van der Ende[22], Sheila Weintraub[4], Tomasz Wolanczyk[23], Bernardine Woo[13], Bahr Weiss[24], John Weisz[25], Rita Zukauskiene[26], Frank C. Verhulst[22]

[1]*Bryn Mawr College*
[2]*Columbia University*
[3]*University of Vermont*
[4]*Helsinki University*
[5]*University of Zagreb*
[6]*University of Southern Denmark*
[7]*Columbia University*
[8]*Babes-Bolyai University*
[9]*Ankara University*
[10]*Oregon Health Sciences University*
[11]*University of Coimbra*
[12]*Scientific Institute, IRCCS Eugenio Medea*
[13]*Institute of Mental Health, Singapore*
[14]*3-C Institute for Social Development*
[15]*Chinese University of Hong Kong*
[16]*Shandong University*
[17]*University of Belgrade*
[18]*University of Novi Sad*
[19]*Tehran Research Institute of Exceptional Children*
[20]*Attiki Child Psychiatric Hospital*
[21]*University of Harran*
[22]*Erasmus University Medical Center-Sophia Children's Hospital*
[23]*Medical University of Warsaw*
[24]*Vanderbilt University*

Correspondence should be addressed to Leslie A. Rescorla, Bryn Mawr College, 101 N. Merion Avenue, Bryn Mawr, PA 19010. E-mail: lrescorl@brynmawr.edu

Color versions of one or more figures in the article can be found online at www.taylorandfrancis.com/hcap.

[25]Harvard University
[26]Mykolas Romeris University

Parent–teacher cross-informant agreement, although usually modest, may provide important clinical information. Using data for 27,962 children from 21 societies, we asked the following: (a) Do parents report more problems than teachers, and does this vary by society, age, gender, or type of problem? (b) Does parent–teacher agreement vary across different problem scales or across societies? (c) How well do parents and teachers in different societies agree on problem item ratings? (d) How much do parent–teacher dyads in different societies vary in within-dyad agreement on problem items? (e) How well do parents and teachers in 21 societies agree on whether the child's problem level exceeds a deviance threshold? We used five methods to test agreement for Child Behavior Checklist (CBCL) and Teacher's Report Form (TRF) ratings. CBCL scores were higher than TRF scores on most scales, but the informant differences varied in magnitude across the societies studied. Cross-informant correlations for problem scale scores varied moderately across societies studied and were significantly higher for Externalizing than Internalizing problems. Parents and teachers tended to rate the same items as low, medium, or high, but within-dyad item agreement varied widely in every society studied. In all societies studied, both parental noncorroboration of teacher-reported deviance and teacher noncorroboration of parent-reported deviance were common. Our findings underscore the importance of obtaining information from parents and teachers when evaluating and treating children, highlight the need to use multiple methods of quantifying cross-informant agreement, and provide comprehensive baselines for patterns of parent–teacher agreement across 21 societies.

Dirks, De Los Reyes, Briggs-Gowan, Cella, and Wakschlag (2012) argued that assessment methods such as rating scales, interviews, and observational instruments "emphasize psychopathology as a trait that will generalize across situations" (p. 560). This assumption of cross-situational generality is challenged by the fact that cross-informant agreement is generally quite modest. For example, meta-analyses by Achenbach, McConaughy, and Howell (1987) yielded a mean correlation (r) of only .27 between parent and teacher reports of children's problems. Subsequent studies have confirmed the meta-analytic findings (De Los Reyes & Kazdin, 2005). Although these robust findings challenge the assumption of cross-situational generality of psychopathology, they have also prompted theorists to consider how incorporating data on cross-informant agreement might improve the diagnostic process for clinical cases. For example, Dirks et al. proposed that cross-informant findings can indicate the pervasiveness of behavioral symptoms across contexts, as illustrated by a study in which boys with conduct problems could be divided into two subgroups, those aggressive in all interpersonal contexts versus those aggressive only in response to peer provocation (Wright & Zakriski, 2001).

De Los Reyes (2011) and Thomas, Goodman, and Kundey (2013) suggested that, rather than being viewed as measurement error, modest cross-informant agreement reveals important variability in behavior across contexts, such as home versus school. Supporting De Los Reyes's argument, Achenbach (2011) cited behavior genetic research indicating that much of the variance in parent and teacher ratings can be accounted for by genetically and environmentally influenced behaviors that differ across contexts. In addition, Achenbach suggested that low agreement may arise because parents and teachers cannot observe behavior in settings where they are not present. Furthermore, transactional effects may limit cross-informant agreement because parents and teachers elicit different child behaviors. For example, De Los Reyes, Henry, Tolan, and Wakschlag (2009) reported that disruptive behavior observed in parent-child lab interactions predicted of parents' reports of disruptive behavior but not teachers' reports. Conversely, disruptive behavior observed in examiner–child lab interactions predicted teachers' reports of disruptive behavior but not parents' reports. Equally important, as noted by De Los Reyes et al. (2013), parents and teachers may have different "decision thresholds" for judging behavior to be problematic. For example, parents may be more likely than teachers to see certain behaviors as problematic because they lack large reference groups with whom to compare their own children.

De Los Reyes et al. (2013) suggested that, when discrepant informant reports appear to reflect measurement error, it may be useful to employ a combinational algorithm such as the "and" rule (i.e., the problem is considered present if both informants say it is) or the "or" rule (e.g., the problem is considered present if

either informant says it is). However, when discrepant reports appear to reflect important contextual influences, both of these algorithms may sacrifice valuable clinical information. Specifically, the "or" rule provides no information about the context in which problematic behaviors occur (e.g., home vs. school), whereas the "and" rule fails to identify cases who display only context-specific problems (e.g., a problem occurs only at home). By examining factors that relate to cross-informant agreement, our study of parent–teacher cross-informant agreement in 21 societies may help to advance knowledge about the advisability of using combinatorial algorithms.

INTERNATIONAL STUDIES OF PARENT–TEACHER AGREEMENT

Major reviews of parent–teacher agreement have involved primarily U.S. samples (Achenbach et al., 1987; De Los Reyes & Kazden, 2005). However, parent–teacher agreement might differ across societies because schools differ across societies. For example, Pong and Pallas (2001) reported large differences in class size and degree of national "centralization" of educational practices, curriculum, and standards across nine societies, ranging from uniformly large classes and strong centralization in Singapore and Hong Kong to more variable but generally smaller class sizes and extreme decentralization in the United States. Cultural differences in parental expectations about children's learning and parental involvement in schools have also been reported (e.g., Phillipson & Phillipson, 2007). Although parent–teacher agreement studies have revealed societal variations, as summarized next, these studies vary in the ages studied, scales analyzed, and statistics used, making it hard to synthesize the results.

Some societal variation has been evident in cross-informant correlations (rs) between ratings obtained via the Child Behavior Checklist (CBCL) and Teacher's Report Form (TRF; Achenbach & Rescorla, 2001). For example, rs ranged from .33 to .50 across age/gender groups for Total Problems scores in Italy (Frigerio et al., 2004), from .15 to .51 across scales in Croatia (Rudan, Begovac, Scirovicza, Filipovic, & Skocic, 2005), and from .13 to .33 for Chinese 6- to 11-year-olds (Deng, Liu, & Roosa, 2004). Similarly, for Total Difficulties on the Strengths and Difficulties Questionnaire (Goodman, 1997), parent–teacher rs of .37 to .62 ($M = .44$) were reported across eight studies conducted in the United Kingdom, Australia, Belgium, the Netherlands, and China (Stone, Otten, Engels, Vermulst, & Janssens, 2010).

Cross-informant agreement has also been studied with respect to mean scale scores. Several studies have reported that CBCL scores are generally higher than TRF scores, with some variation across societies. For example, Deng et al. (2004) reported higher CBCL than TRF scores in China for the three scales they studied (Attention Problems, Internalizing, and Externalizing), and Rudan et al. (2005) reported that CBCL scores were higher than TRF scores in Croatia on every narrow-band scale except Attention Problems. Van der Ende, Verhulst and Tiemeier (2012) reported that CBCL scores were higher than TRF scores in the Netherlands but more so for Externalizing than for Internalizing.

Dichotomous agreement based on scores above or below a deviance cutpoint also appears to differ somewhat across societies, but this is harder to determine as studies have varied in statistics reported, such as odds ratios (ORs), kappas, or decision statistics. Studies have also differed in the deviance cutpoint used, namely, the "positive test" probability (Kraemer, 1992). For example, using scores greater than the 90th percentile for Chinese 6- to 11-year-olds as the deviance cutpoint, Deng et al. (2004) reported kappas of .11 to .24 for parent–teacher agreement, suggesting modest agreement. Using scores greater than the 98th percentile as the deviance cutpoint, Liu et al. (2001) reported an OR of 18.9 for Chinese adolescents scoring in the deviant range on the TRF given a score in the deviant range on the CBCL, suggesting quite good agreement. When Goodman, Ford, Simmons, Gatward, and Meltzer (2000) used a Strengths and Difficulties Questionnaire criterion of "probable" as the deviance cutpoint (10.5% of their British sample), they obtained sensitivities of 30 to 34% for parents and 35 to 39% for teachers with respect to diagnoses. However, sensitivities increased to 59 to 62% when an "or" rule was used (i.e., deviance according to either parent or teacher ratings), suggesting that parents and teachers each contribute valid but different information about clinical deviance.

To our knowledge, no study has systematically compared parent versus teacher reports across multiple societies using the same instruments and analyses, an approach that is needed to determine the degree of societal variation in cross-informant agreement. However, Rescorla et al. (2013) compared agreement between parents' ratings on the CBCL and self-ratings on the Youth Self-Report (YSR; Achenbach & Rescorla, 2001) for 27,861 dyads in 25 societies. Mean cross-informant rs varied considerably across societies studied (from .17 to .58, $M = .41$). YSR scores were significantly higher than CBCL scores in every society studied, but the informant effect size (ES) varied significantly across societies and problem scales. When Q correlations between mean item ratings derived from all parents and all adolescents in each society were computed, the results (range = .72–.94) indicated that, on average, parents and adolescents in every society tended to rate the same problems as low, medium, or high. However, Q correlations for within-dyad parent–adolescent

agreement varied widely within each society, yielding modest mean dyadic Qs in all societies. With scores greater than the 84% percentile as the deviance cutpoint, most parent–adolescent dyads agreed on the adolescents' deviance status (71% to 85% across societies). Disagreements were about evenly divided between parental noncorroboration of self-reported deviance and adolescent noncorroboration of parent-reported deviance.

PURPOSE OF THE CURRENT STUDY

The increasing diversity of families served by today's clinicians makes it especially important to document aspects of parent–teacher agreement that transcend single societies. To expand the evidence regarding parent–teacher agreement, the fundamental purpose of this study was to answer questions about societal variations in parent–teacher agreement by using samples from many societies. We therefore analyzed parent and teacher ratings for 27,962 children from 21 societies to answer the following questions: (a) Do parents report more problems than teachers, and does this vary by society, age, gender, and type of problem? Based on previous research, we hypothesized that parents would report more problems than teachers, that the informant effect would vary across the societies and types of problems studied, and that age and gender effects would be small. (b) Do parent–teacher cross-informant rs vary between different problem scales and between societies? Based on previous research, we hypothesized that cross-informant rs would be larger for externalizing than for internalizing problems and would vary across the societies studied. (c) How well do parents and teachers in different societies agree on which problem items they tend to rate low, medium, or high? We predicted that Q correlations between parent and teacher mean item ratings would be large in all the societies studied. (d) How much do parent–teacher dyads in different societies vary in within-dyad agreement on problem items? We predicted that within-dyad Q correlations would vary widely in every society but be modest on average in each society studied. (e) How well do parents and teachers in different societies agree on whether a child has enough problems to exceed a deviance threshold? We hypothesized that categorical dyadic agreement would be modest in all the societies studied, with both parents and teachers failing to corroborate reports of deviance by the other informant in many cases. It should be noted that although we predicted variation across societies for all our analyses, too little previous evidence was available to justify hypotheses regarding which of the societies studied would have lower versus higher levels of agreement. We therefore regarded this aspect of our study as exploratory.

METHOD

CBCL and TRF ratings were collected by indigenous investigators (Table 1) and sent to the lead authors for analysis. Previous reports for a few samples presented parent–teacher agreement results, but the total sample sizes, items, and scales analyzed differed from those we analyzed. Participants in the CBCL–YSR study (Rescorla et al., 2013) were included if TRF ratings were also available.

Participants

Participants were recruited through schools or household registers/addresses. CBCL data were collected by interviews, by mail, or via CBCLs brought home from school, whereas TRFs were completed at school. Samples ranged from 357 to 4,444 ($N = 27,962$ parent–teacher dyads; 50% boys). To have sufficient cell sizes for analyzing effects of gender, society, and informant within each age bracket, we required 70 or more per age bracket, consistent with Rescorla et al. (2013). We restricted our age range to 6 to 16 years because few samples included ages older than 16 years. For correlational analyses, we used data from all societies for ages 6 to 16. For analyses of variance (ANOVAs), all 21 societies were used in analyses for ages 6 to 8 and 9 to 11 ($n = 17,933$), whereas 14 societies were used in analyses for ages 6 to 8, 9 to 11, 12 to 14, and 15 to 16 ($n = 22,192$). Children referred for mental health services had been excluded from the samples we received from seven societies. Investigators followed conventions of their institutions regarding informed consent.

Measures

Parents and teachers completed translated versions of the instruments in 19 societies and English-language versions in the United States and Jamaica. Translations approximated the fifth-grade reading level of the U.S. versions, and independent back-translations were done. CBCL and TRF problem items are rated as 0 (*not true* [*as far as you know*]), 1 (*somewhat or sometimes true*), or 2 (*very true or often true*). When the CBCL and TRF were revised in 2001, six items were replaced on the CBCL, three of which were also replaced on the TRF. Because some samples used pre-2001 versions of the forms, we omitted these six items, as well as items not shared by the forms, plus two open-ended items, leaving 90 items common to the CBCL and TRF.

Achenbach and Rescorla's (2001) exploratory and confirmatory factor analyses of the CBCL and TRF yielded eight empirically based syndromes: Anxious/Depressed, Withdrawn/Depressed, and Somatic Complaints (all loading on the broad-band, second-order Internalizing factor); Rule-Breaking Behavior and Aggressive Behavior (both loading on the broad-band,

TABLE 1
Reference, Total Sample Size, Ages, Sampling Procedure, Percentage Male,
Response Rate, and Exclusion of Referred Children for 21 Societies

Society	Reference	N	Ages	Sampling Procedure	% Male	Response Rate	Referred Excluded
China	Liu et al., 1999	4,444	6–16	regional household-based	51%	92%	No
Croatia	Rudan et al., 2005	1,920	6–16	national school-based	48%	98%	No
Denmark	Bilenberg, 1999	602	6–16	regional household-based	48%	84%	No
Finland	Weintraub et al., 2004	1,635	6–16	regional school-based	49%	92%	No
France	Fombonne & Vermeersch, 1997	482	6–11	national sample of utility employees	48%	unknown	Yes
Greece	Roussos et al., 1999	1,157	6–12	national school-based	49%	99%	No
Hong Kong	Leung et al., 2006	1,560	6–15	city school-based	49%	83%	No
Iran	Minaei, 2005	1,096	6–15	regional school-based	55%	100%	Yes
Italy	Frigerio et al., 2004	917	6–15	regional school-based	49%	79%	Yes
Jamaica	Lambert, Lyubansky, & Achenbach, 1999	660	6–15	national school-based	50%	87%	Yes
Lithuania	Zukauskiene et al., 2004	2,450	6–15	national school-based sample	48%	84%	No
Netherlands	Tick et al., 2007	813	6–16	regional household	51%	74%	No
Poland	Wolanczyk, 2003	2,454	6–15	national school-based	49%	99%	No
Portugal	Fonseca et al., 1995	1,373	6–15	regional school-based	50%	91%	No
Puerto Rico	Achenbach et al., 1990	614	6–16	island-wide household-based	50%	93%	No
Romania	Domuta, 2004	1,250	6–16	regional household-based	49%	80%	Yes
Serbia	Markovic et al. 2011	484	6–11	regional school-based	50%	70%	No
Singapore	Woo et al., 2007	1,535	6–11	national school-based	49%	60%	No
Thailand	Weisz et al., 1987, 1989	357	6–11	national school-based sample	50%	92%	Yes
Turkey	Erol & Simsek, 1997	1,507	6–16	national household-based	53%	88%	Yes
U.S.	Achenbach et al., 2001	872	6–16	national household-based	52%	88%	No

Note: Published sources for data sets are listed in the references. All other sources are unpublished.

second-order Externalizing factor); and Social Problems, Thought Problems, and Attention Problems (not loading differentially on either second-order factor). Using the 90 items in common, we calculated 2001 syndrome scale scores for both forms. Internalizing and Externalizing scales were scored by summing the 0-1-2 ratings of the items on their constituent syndromes, whereas the broad-band Total Problems scale was scored by summing the ratings of all 90 problem items analyzed for this study. We also analyzed six scales comprised of items identified by experts from 16 societies as very consistent with diagnostic categories of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-IV*; American Psychiatric Association, 1994). These *DSM*-oriented scales included Anxiety Problems, Affective Problems, Somatic Problems, Attention Deficit Hyperactivity Problems, Oppositional Defiant Problems, and Conduct Problems. Achenbach and Rescorla (2001) reported alphas of .90 to .97 for the broad-band scales on both the CBCL and TRF; alphas for the syndrome and *DSM*-oriented scales ranged from .72 to .94 on the CBCL and from .72 to .95 on the TRF.

Overview of Data Analyses

CBCL and TRF problem scale scores were positively skewed in every sample because relatively few problems are reported for most individuals in population samples. However, because general linear models are very robust with respect to deviations from normality, especially with very stringent criteria for significance and large samples having similar skew (Kirk, 1995), we analyzed untransformed raw scale scores.

We first performed ANOVAs on CBCL and TRF scale scores. To take account of the fact that CBCL and TRF ratings comprise correlated observations, we used repeated measures (mixed-model) ANOVA for each of the 17 problem scales, with informant (parent vs. teacher) as the within-subjects factor and gender, age bracket, and society as between-subjects factors. All interactions were tested. We treated society as a fixed effect because we do not attempt to generalize our results beyond the societies included in the analyses. These ANOVAs were run first for ages 6 to 11 across 21 societies and then for ages 6 to 16 across 14 societies. Significant ESs, reported as partial η^2, were interpreted using Cohen's (1988) criteria (*small* = 1–5.9%, *medium* = 6–13.9%, and *large* $\geq$ 14%). As an alternative approach to analyzing informant score differences with correlated observations, we obtained generalized estimating equations (GEE; Hanley, Negassa, Edwards, & Forrester, 2003) for Total Problems scores, using the same four factors of informant, gender, age bracket, and society. GEE is useful for non-normal data and provides more precise estimates of standard errors than GLM.

For our second set of analyses, we computed cross-informant rs for each of the 17 scale scores within each society. For our third set of analyses, we computed a Q

correlation between the 90 mean CBCL item ratings and the 90 mean TRF item ratings within each society, based on the procedures pioneered by Stephenson (1935), whereby Pearson correlations are calculated between informants across items. For our fourth set of analyses, we computed Q correlations between CBCL and TRF item ratings within each parent–teacher dyad, converted these to Fisher's zs, and then submitted them to an ANOVA to test effects of society, age, and gender on the dyadic correlations. For our fifth set of analyses, we analyzed parent–teacher agreement regarding deviance status within each society on Internalizing and Externalizing. To use all 21 societies, we conducted these analyses for ages 6 to 11 only. For each society, we set as our deviance cutpoint the score at 1 SD above its mean (i.e., the "test positive" percentage $= 16\%$). Where within-society ANOVAs indicated significant age and/or gender effects, we assigned cutpoints separately for each age and/or gender group.

In view of the high statistical power afforded by our large sample size, we present only the findings that were significant at $p < .001$. Because so many findings reached this stringent alpha level, we present ESs, rather than individual F and p values. Even with this stringent alpha level, power was 95% to detect ESs of $\eta^2 \geq .0025$ (equivalent to $f \geq .045$; Faul, Erdfelder, Buchner, & Lang, 2009).

RESULTS

Parent–Teacher Differences in Mean Scale Scores

Table 2 contains the CBCL and TRF *omnicultural means* (average of the 21 society means) for all 17 scales based on the full sample from each society. CBCL omnicultural means were higher than TRF omnicultural means for all scales except Withdrawn/Depressed. The standard deviation of the omnicultural mean for each scale was much smaller than the within-society standard deviation, indicating more variance within than between societies. As depicted in Figure 1, the mean CBCL Total Problems score was higher than the mean TRF Total Problems score in all societies except Thailand. However, the informant difference was larger in some societies (e.g., Singapore) than in others (e.g., Jamaica).

For the repeated measures ANOVAs on each problem scale for ages 6 to 11 across 21 societies, Table 3 presents ESs for all four main effects and all 11 interactions. Interaction effects that were not significant or had ESs <1% are not described below. Main effects for ages 6 to 16 are described next, as well as any interactions with ESs >1%.

For Total Problems scores, the ANOVA yielded a significant effect of informant (ES = 10%), a very small gender effect (ES = 1%, boys > girls), a nonsignificant

TABLE 2
Omnicultural Means and Standard Deviations for Broad-Band,
Syndrome, and *DSM*-Oriented Scales Across 21 Societies ($N = 27,962$)

Scale	CBCL Omnicultural M (SD)	CBCL Omnicultural SD	TRF Omnicultural M (SD)	TRF Omnicultural SD
Broad-Band Scales				
Total Problems (90)[a]	21.1 (4.2)	15.5	14.6 (4.6)	14.8
Internalizing Problems (29)	6.4 (1.6)	5.4	4.7 (1.7)	5.0
Externalizing Problems (26)	6.2 (1.3)	5.5	3.9 (1.4)	5.8
Syndromes				
Anxious/Depressed (13)	3.4 (0.9)	3.0	2.3 (0.9)	2.6
Withdrawn/Depressed (7)	1.8 (0.5)	1.9	1.8 (0.7)	2.3
Somatic Complaints (9)	1.3 (0.4)	1.8	0.6 (0.3)	1.3
Social Problems (11)	2.7 (0.6)	2.6	1.7 (0.6)	2.4
Thought Problems (10)	1.4 (1.6)	1.5	0.6 (0.3)	1.2
Attention Problems (12)	4.2 (0.8)	3.5	3.5 (1.0)	3.8
Rule-Breaking Behavior (9)	1.5 (0.4)	1.7	1.1 (0.4)	1.8
Aggressive Behavior (17)	4.8 (1.0)	4.3	2.9 (1.0)	4.4
DSM-Oriented Scales				
DSM–Affective Problems (8)	1.1 (0.3)	1.5	0.9 (0.4)	1.5
DSM–Anxiety Disorder (6)	2.1 (1.3)	1.7	1.1 (0.4)	1.5
DSM–Somatic Problems (7)	1.0 (0.3)	1.4	0.4 (0.2)	1.0
DSM–Attention Deficit (5)	2.4 (0.5)	2.1	1.7 (0.5)	2.2
DSM–Oppositional Defiant (4)	1.8 (0.4)	1.6	0.9 (0.3)	1.4
DSM–Conduct Problems (11)	2.0 (0.4)	2.0	1.3 (0.5)	2.5

Note: *DSM* = Diagnostic and Statistical Manual of Mental Disorders; CBCL = Child Behavior Checklist for Ages 6–18 (Achenbach & Rescorla, 2001); TRF = Teacher's Report Form (Achenbach & Rescorla, 2001). Omnicultural mean (*SD*) = mean of 21 society means and the standard deviation of that mean; omnicultural *SD* = mean of 21 society standard deviations.
[a]Parenthetical numerals indicate the number of items comprising each scale, based on 90 items analyzed.
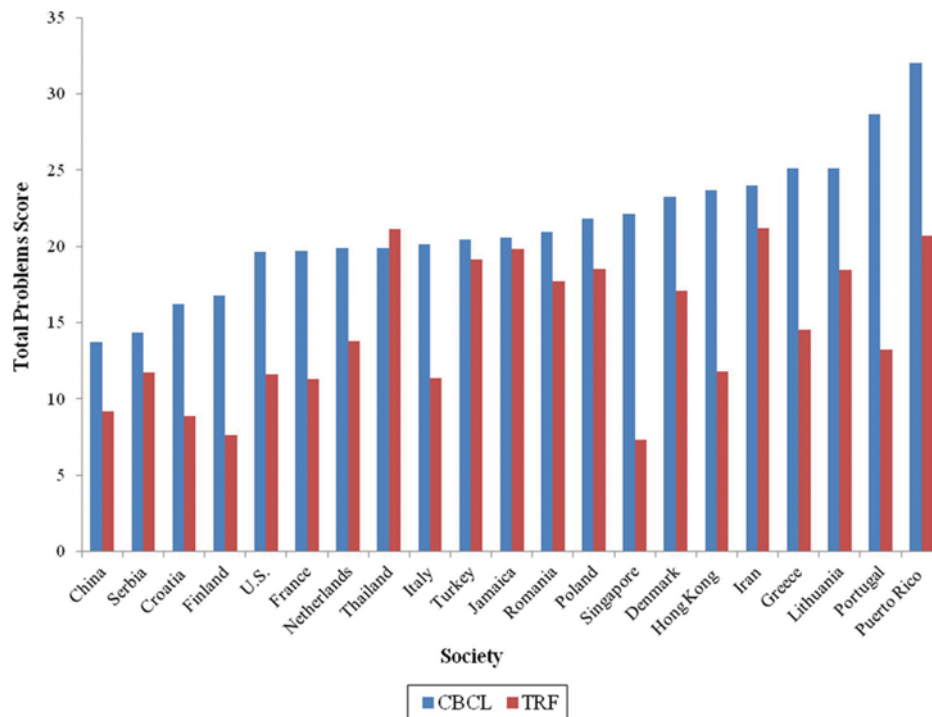
FIGURE 1   Child Behavior Checklist (CBCL) and Teacher's Report Form (TRF) Mean Total Problems Scores in 21 Societies.

age effect, and a significant effect of society (ES = 10%). The significant Informant × Society interaction (ES = 5%) indicated that the size of the CBCL–TRF discrepancy varied somewhat across societies. Results of the ANOVA for ages 6 to 16 in 14 societies yielded a slightly smaller informant ES (6%), but the other three ESs were quite similar: gender <1% (boys > girls), age <1% (younger < older), society = 10%, Informant × Society interaction = 3%, and all other ESs ≤1% or not significant.

TABLE 3
Percentage of Variance Accounted for by Analysis of Variance Effects (η^2) for Ages 6 to 11 for 21 Societies ($n = 17,933$)

Scale	I	G	A	S	I × G	I × A	I × S	IGA	IGS	IAS	4-Way	G × A	G × S	A × S	GAS
Broadband Scales															
Total Problems	10	1	—	10	<.1	—	5	—	—	—	—	—	<1	<1	—
Internalizing Problems	4	<1	<1	13	—	—	4	—	—	—	—	—	—	—	—
Externalizing Problems	10	3	—	6	<1	<1	4	—	—	<1	—	—	<1	<1	—
Syndromes															
Anxious/Depressed	6	<1	<1	18	<1	—	4	—	—	—	—	—	—	—	—
Withdrawn/Depressed	<1[a]	—	<1	8	—	—	3	—	—	—	<1	—	—	—	—
Somatic Complaints	7	<1	<1	3	<1	<1	3	—	—	—	—	—	—	<1	—
Social Problems	10	<1	—	6	<1	<1	4	—	—	—	—	—	<1	—	—
Thought Problem	5	<1	—	7	<1	—	4	—	—	—	—	—	<1	<1	<1
Attention Problems	3	4	—	8	<1	—	4	—	—	<1	—	—	<1	<1	—
Rule-Breaking Behavior	5	3	—	5	<1	<1	3	—	—	—	—	—	<1	<1	—
Aggressive Behavior	10	3	—	6	<1	<1	4	—	—	<1	—	—	<1	<1	—
DSM-Oriented Scales															
DSM-Affective Problems	1	ns	<1	7	<1	—	3	—	—	—	—	—	<1	<1	—
DSM-Anxiety Disorder	8	—	—	12	—	—	4	—	—	—	—	—	—	—	—
DSM-Somatic Problems	7	<1	<1	3	—	—	3	—	—	—	—	—	—	<1	—
DSM-Attention Deficit	7	4	<1	5	<1	<1	3	—	—	<1	—	—	<1	<1	—
DSM-Oppositional Defiant	17	2	—	4	<1	<1	5	—	—	—	—	—	<1	<1	—
DSM-Conduct Problems	<1	4	—	4	<1	—	2	—	—	<1	—	—	<1	<1	—

Note: All effect sizes reported were significant at the $p < .001$ level; blank cells were not significant at $p < .001$. I = informant; G = gender; A = age group (6–8, 9–11), S = society; DSM = *Diagnostic and Statistical Manual of Mental Disorders*.
[a]TRF > CBCL for Withdrawn/Depressed.

The ANOVAs for ages 6 to 11 and 6 to 16 for Internalizing both yielded informant ESs of 4% (CBCL > TRF), gender ESs <1% (girls > boys), and age ESs ≤1% (younger < older). The societal ES was 13% for ages 6 to 11 and 12% for ages 6 to 16, with Informant × Society interactions of 4% and 2%, respectively (large CBCL > TRF effects in Portugal and Finland; small effects in Jamaica and Serbia). For Externalizing, the informant ES was 10% for ages 6 to 11 and 6% for ages 6 to 16 (CBCL > TRF). For ages 6 to 11 only, the informant ES was larger for Externalizing (10% = lower agreement) than for Internalizing (4% = higher agreement). For Externalizing, the ES for gender was 3% for ages 6 to 11 and 1% for ages 6 to 16 (boys > girls). The age effect was not significant in either ANOVA. The societal ES was 6% for both ANOVAs, with Informant × Society interactions of 4% (6–11) and 2% (6–16). Externalizing scores were generally higher on the CBCL than the TRF, with large differences in Portugal and Puerto Rico and minimal differences in Jamaica and Romania.

For the syndromes and *DSM*-oriented scales, ESs for informant ranged from <1% to 17% for ages 6 to 11 and from <1% to 12% for ages 6 to 16, with CBCL means > TRF means for all scales except Withdrawn/Depressed, where the pattern was reversed. For ages 6 to 11, gender effects were not significant or ≤1% for eight scales, and 2% to 4% for the six scales tapping attention and externalizing kinds of problems (boys > girls). For ages 6 to 16, gender ESs >1% (boys > girls) were found only on Attention Problems (3%), Rule-Breaking Behavior (2%), *DSM*-Attention Deficit Hyperactivity (2%), and *DSM*-Conduct Problems (3%). Age ESs were either not significant or ≤1%. Societal ESs for ages 6 to 11 ranged from 3% (Somatic Complaints, *DSM*-Somatic Problems) to 18% (Anxious/Depressed). For ages 6 to 16, societal ESs ranged from 3% for *DSM*-Somatic Problems to 16% for Anxious/Depressed. Informant × Society interactions ranged from 2% to 5% for ages 6 to 11 and 1% to 3% for ages 6 to 16, indicating modest variation across societies in agreement. For ages 6 to 16, all remaining interactions were not significant or ≤1%, except for a 2% Age × Society interaction ES for Thought Problems.

The supplementary GEE analysis was conducted for Total Problems for the 14 societies with data for ages 6 to 16 (*n* = 22,192). Predictors were informant, gender, age group, and society, plus all 11 interactions. We used an identity link matrix with an unstructured correlation matrix. Results indicated significant Wald chi-square values (*p* < .001) for all four main effects and for many of the interactions. The unstandardized betas for Total Problems (which has a range from 0 to 180 based on the 90 items we analyzed) were 8.57 points for informant (CBCL > TRF) and 3.2 points for gender (boys > girls). Betas for age contrasts ranged from −2.98 to 1.79 (ages 11–14 > ages 15–16 > ages 9–11 > ages 6–8). Betas for

society ranged from −3.30 to 12.38, and betas for the Informant × Society interaction ranged from −9.18 to 4.48 points. Details of these results are available from the first author upon request.

Parent–Teacher Correlations for Scale Scores

Table 4 presents the range of mean cross-informant *r*s for all 17 scales across societies and each scale's omnicultural mean *r* (mean of the *r*s for the 21 societies). The largest omnicultural cross-informant *r*s were for the six scales tapping attention and externalizing kinds of problems, ranging from .28 (*DSM*-Oppositional Defiant) to .37 (Attention Problems). Cross-informant agreement was not higher for scales with many rarely endorsed items than for those with few rarely endorsed items. To illustrate this, we used the asymptotic variance *z* test to compare the cross-informant *r* for Thought Problems (many rarely endorsed items) with that for Attention Problems (few rarely endorsed items) for the full sample *r*s of .18 and .38 (slightly higher than the omnicultural *r*s of .17 and .37). The *z* of 29.96 (*p* < .001) indicated that cross-informant agreement was much higher for Attention Problems than for Thought Problems.

The omnicultural cross-informant *r* was higher for Externalizing (*r* = .32) than for Internalizing (*r* = .21). To test the significance of this difference, we used the asymptotic variance *z* test on the cross-informant *r*s for the full sample (.33 for Externalizing and .24 for

TABLE 4
Cross-Informant Scale Score Correlations
for Full Sample (*N* = 27,962)

Scale	Minimum	Maximum	Omnicultural Mean
Broad-Band Scales			
Total Problems	.07	.60	.29
Internalizing	.00	.44	.21
Externalizing	.07	.59	.32
Syndromes			
Anxious/Depressed	−.04	.39	.17
Withdrawn/Depressed	.11	.43	.24
Somatic Complaints	.03	.48	.19
Social Problems	.09	.62	.26
Thought Problems	.04	.48	.16
Attention Problems	.19	.62	.37
Rule Breaking Behavior	.09	.53	.29
Aggressive Behavior	.07	.57	.30
DSM-Oriented Scales			
DSM–Affective Problems	.04	.50	.20
DSM–Anxiety Problems	.01	.43	.18
DSM–Somatic Problems	.06	.48	.19
DSM–Attention Deficit	.17	.60	.35
DSM–Oppositional Defiant	.09	.50	.28
DSM–Conduct Problems	.15	.59	.34

Note: *DSM* = *Diagnostic and Statistical Manual of Mental Disorders*.

Internalizing, slightly higher than the omnicultural rs). Results indicated that the Externalizing r was significantly higher ($z = 13.1$, $p < .001$).

We also calculated cross-informant rs for the full sample separately by gender for the three broad-band scales. The cross-informant r for Total Problems did not differ significantly by gender (.31 for boys vs. .29 for girls). However, gender differences for Internalizing (.22 for boys vs. .26 for girls) and for Externalizing (.34 for boys vs. .29 for girls) were both significant at $p < .001$ by Fisher's z test for independent samples, suggesting a Gender × Problem Type interaction.

When we averaged the rs for each of the 17 problem scales within each society, mean societal cross-informant rs ranged from .09 (Thailand) to .49 (Denmark), with an omnicultural mean $r = .26$. We applied the between-subjects Fisher's z test to these societal differences in mean cross-informant rs, starting with the most extreme values and proceeding toward the middle in pairs. Only the first five comparisons were significant at $p < .001$, after which the comparisons were not significant.

Cross-Informant Agreement on Item Ratings

Within each society, the Q correlations between the 90 CBCL mean item ratings and the 90 TRF mean item ratings were large, ranging from .65 (United States) to .90 (Lithuania), with an omnicultural mean of .77 (see Table 5). These results indicate strong within-society agreement between parents and teachers, on average, regarding which items received low, medium, or high ratings, as well as considerable across-society similarity in levels of parent–teacher mean item agreement.

When we identified the 30 items with the highest omnicultural mean ratings on the CBCL and the TRF, 24 items were common to both "top 30" lists (see Table 6). The CBCL omnicultural mean item rating significantly exceeded the TRF omnicultural mean item rating for 21 of these 24 items. Societal ESs ranged from 3% (1. Argues a lot) to 22% (32. Feels he/she has to be perfect). The CBCL base rate (ratings of 1 or 2) significantly exceeded the TRF base rate for all 24 items, but base rate differences varied across items. Half of the 24 "top" items came from the Attention Problems and Aggressive Behavior syndromes, with the rest drawn from the Anxious/Depressed, Withdrawn/Depressed, Social Problems, and Rule-Breaking Behavior syndromes or from Other Problems (items not loading on a syndrome).

Dyadic Q Correlations for Item Ratings

When we calculated Q correlations between the 90 item ratings by each parent and teacher dyad, within-society mean dyadic Qs ranged from .17 for Thailand to .30 for Lithuania (omnicultural $M = .23$), with substantial

TABLE 5
Scale and Item Ratings: Correlations by
Society for Full Sample ($N = 27,962$)

Society	CBCL × TRF Mean Scale r	CBCL × TRF Mean Item Q^a	CBCL × TRF Mean Dyadic Q (SD)^b
Thailand	.09	.71	.17 (.15)
Singapore	.15	.69	.18 (.18)
Hong Kong	.16	.74	.18 (.17)
Turkey	.16	.77	.20 (.18)
China	.18	.87	.23 (.22)
Iran	.20	.89	.22 (.19)
U.S.	.23	.65	.18 (.18)
Italy	.23	.72	.23 (.18)
Greece	.23	.74	.23 (.18)
Poland	.24	.77	.22 (.18)
Jamaica	.24	.78	.20 (.19)
France	.26	.66	.24 (.19)
Netherlands	.29	.77	.22 (.17)
Portugal	.30	.78	.22 (.17)
Croatia	.30	.79	.24 (.21)
Puerto Rico	.31	.66	.22 (.17)
Finland	.31	.69	.19 (.19)
Lithuania	.31	.90	.30 (.20)
Serbia	.35	.87	.29 (.23)
Romania	.36	.85	.28 (.27)
Denmark	.49	.83	.29 (.19)
M^c	.26	.77	.23 (.04)

Note: The table is arranged in ascending order by CBCL × TRF mean scale r, which is the average within each society of the 17 scale rs. CBCL = Child Behavior Checklist for Ages 6–18 (Achenbach & Rescorla, 2001); TRF = Teacher's Report Form (Achenbach & Rescorla, 2001).

^{a}The mean item Q for each society was calculated by correlating the mean item ratings for all parents with the mean item ratings for all teachers.

^{b}The mean dyadic Q for each society was calculated by averaging across all dyads in the society's sample the dyadic Qs derived from correlating the 90 item ratings for each parent-teacher dyad in the sample. The dyadic SD for each society was the SD of the mean dyadic Q.

^{c}The mean for each measure represents the omnicultural mean (i.e., the mean of the 21 society means).

within-society variation across dyads in item agreement. Dyadic Qs were much lower than the Qs between mean item ratings (see Table 5), indicating greater variance within than between societies on item ratings. When dyadic Qs converted to Fisher's zs were submitted to a 21 (society) × 2 (gender) × 2 (age) ANOVA, dyadic Qs were slightly higher for boys than for girls (mean $Q = .24$ vs. .22, ES < 1%), but they did not vary by age. The societal ES was 3% and the Society × Age interaction was <1%, indicating minor variation across societies in mean dyadic agreement.

Within-society mean dyadic Qs for Internalizing items ranged from .16 (Thailand and Hong Kong) to .32 (Lithuania), with an omnicultural mean of .23 and ESs of 2% for society and for the Society × Age interaction. Similar results were obtained for Externalizing Q correlations, which ranged from .17 (Turkey) to .38 (Serbia),

TABLE 6
The 24 Problems in Common from the Child Behavior Checklist (CBCL) and
Teacher's Report Form (TRF) Items with the 30 Highest Mean Ratings

Item	Scale	CBCL M	TRF M	ESs[a] I, S, I × S	CBCL[b] Base Rate	TRF[c] Base Rate
3. Argues a lot	AGG	.78	.36	14%, 5%, 5%	62%	30%
10. Can't sit still, restless or hyperactive	ATT	.61	.37	6%, 7%, 3%	45%	28%
19. Demands a lot of attention	AGG	.59	.36	6%, 12%, 3%	44%	26%
32. Feels he/she has to be perfect	A/D	.53	.39	2%, 22%, 2%	40%	30%
8. Has trouble concentrating/paying attention	ATT	.53	.50	<1%, 4%, 4%	44%	39%
93. Talks too much	OTH	.51	.35	3%, 4%, 2%	39%	26%
63. Prefers being with older kids	OTH	.49	.22	8%, 7%, 2%	40%	18%
86. Stubborn, sullen, or irritable	AGG	.46	.21	9%, 9%, 4%	36%	16%
75. Too shy or timid	W/D	.46	.44	ns, 5%, 2%	37%	35%
11. Clings to adults or too dependent	SOC	.44	.23	5%, 7%, 4%	35%	18%
71. Self-conscious or easily embarrassed	A/D	.42	.38	<1%, 16%, 3%	32%	30%
41. Impulsive or acts without thinking	ATT	.41	.29	2%, 5%, 3%	44%	22%
69. Secretive or keeps things to self	W/D	.40	.39	ns, 15%, 2%	31%	30%
45. Nervous, highstrung, or tense	A/D	.38	.26	2%, 14%, 4%	31%	20%
88. Sulks a lot	AGG	.37	.21	4%, 11%, 3%	28%	16%
17. Daydreams, or gets lost in thoughts	ATT	.36	.35	ns, 13%, 4%	30%	28%
74. Showing off or clowning	OTH	.36	.22	3%, 7%, 3%	27%	16%
87. Sudden changes in mood or feelings	AGG	.33	.20	3%, 5%, 1%	21%	16%
7. Bragging, boasting	OTH	.33	.20	3%, 5%, 2%	28%	16%
26. Doesn't seem to feel guilty after misbehaving	RBB	.33	.24	1%, 4%, 2%	28%	19%
1. Acts too young for his/her age	ATT	.33	.27	<1%, 3%, <1%	27%	20%
61. Poor school work	ATT	.31	.38	<1%, 6%, 2%	26%	29%
42. Would rather be alone than with others	W/D	.30	.23	1%. 6%, 2%	33%	18%
94. Teases a lot	AGG	.30	.20	2%, 9%, 3%	23%	15%

Note: The 24 items in common from the "top 30" lists for both the CBCL and TRF are arranged in descending order by CBCL omnicultural mean item rating. A/D = Anxious/Depressed; W/D = Withdrawn/Depressed; SOC = Social Problems; ATT = Attention Problems; RBB = Rule-Breaking Behavior; AGG = Aggressive Behavior.

[a]ES = effect size from 2 (informant) × 21 (society) mixed-model analysis of variance; I = informant; S = society.

[b]CBCL base rate = percentage of cases rated 1 or 2 on CBCL for full sample of 27,962.

[c]TRF base rate = percentage of cases rated 1 or 2 on TRF for full sample of 27,962.

with an omnicultural mean of .25, a societal ES of 2%, and age and Society × Age interaction (ESs ≤1%).

Cross-Informant Agreement on Deviance

Last, we tested within-dyad dichotomous agreement between a deviant (vs. nondeviant) score on the CBCL and a deviant (vs. nondeviant) score on the TRF for both the Internalizing and Externalizing scales. Deviance was defined as a score >1 *SD* above the mean of the CBCL and TRF scores for each society (i.e., "test positive" percentage of 16%, using age- and/or gender-specific cutpoints where needed).

Table 7 presents the Internalizing and Externalizing decision statistics by society, denoted by their traditional names (e.g., sensitivity, specificity, etc.) and based on parent report as the "outcome" and teacher report as the "test," albeit recognizing that this choice was arbitrary (i.e., teacher report could have been the "outcome"). However, in summarizing our results, we describe the decision outcomes in addition to labeling them with these traditional names.

With agreement defined as both CBCL and TRF scores above the deviance cutpoint or both CBCL and TRF scores below the deviance cutpoint, agreement ranged from 71% (Singapore) to 81% (Denmark) for Internalizing (consistent classification omnicultural *M* = 75%) and from 72% (Thailand) to 81% (Denmark, Lithuania, and Jamaica) for Externalizing (consistent classification omnicultural *M* = 77%). In all societies studied, most children who scored as nondeviant on the CBCL also scored as nondeviant on the TRF: 81% to 89% for Internalizing (specificity omnicultural *M* = 85%) and 82% to 88% for Externalizing (specificity omnicultural *M* = 86%). Furthermore, in all societies studied, most children who scored as nondeviant on the TRF also scored as nondeviant on the CBCL: 82% to 88% for Internalizing (negative predictive value omnicultural *M* = 85%) and 82% to 88% for Externalizing (negative predictive value omnicultural *M* = 86%).

In contrast, a score in the deviant range on the CBCL was corroborated by a score in the deviant range on the TRF for only about one third of the dyads having a deviant CBCL score, with a range from 20% to 48% for Internalizing (sensitivity omnicultural *M* = 30%) and from 19% to 47% for Externalizing (sensitivity omnicultural *M* = 34%). In addition, a score in the deviant range

TABLE 7
Decision Statistics for Ages 6 to 11 by Society for Internalizing/Externalizing ($N = 17,933$)

Scale	SENS	PPV	SPEC	NPV	CC	Kappa
China	.28/.31[a]	.26/.35	.83/.86	.84/.84	.73/.75	.11/.18
Croatia	.28/.38	.29/.40	.86/.87	.86/.86	.76/.78	.14/.25
Denmark	.48/.40	.44/.40	.88/.88	.89/.88	.81/.81	.35/.28
Finland	.25/.37	.27/.31	.85/.82	.84/.86	.74/.74	.10/.17
France	.38/.33	.33/.35	.83/.85	.82/.84	.72/.74	.22/.18
Greece	.22/.31	.25/.30	.85/.86	.83/.87	.74/.77	.07/.17
Hong Kong	.26/.32	.26/.30	.83/.86	.82/.87	.72/.77	.09/.17
Iran	.26/.35	.26/.34	.85/.86	.85/.87	.75/.78	.11/.22
Italy	.31/.28	.32/.34	.84/.86	.84/.82	.74/.78	.16/.15
Jamaica	.24/.47	.30/.47	.86/.88	.82/.88	.74/.81	.11/.35
Lithuania	.37/.43	.35/.43	.84/.88	.85/.88	.75/.81	.21/.31
Netherlands	.29/.29	.29/.29	.86/.88	.86/.86	.77/.77	.16/.26
Poland	.29/.37	.30/.33	.85/.85	.84/.87	.75/.77	.14/.21
Portugal	.33/.35	.33/.37	.85/.88	.86/.87	.76/.79	.19/.23
Puerto Rico	.35/.39	.31/.38	.86/.87	.87/.88	.78/.79	.19/.26
Romania	.32/.43	.30/.43	.85/.87	.86/.88	.76/.79	.17/.30
Serbia	.34/.37	.43/.39	.89/.87	.84/.86	.72/.78	.24/.24
Singapore	.22/.23	.20/.25	.81/.85	.83/.84	.71/.74	.03/.08
Thailand	.20/.19	.21/.20	.84/.87	.82/.86	.72/.72	.03/.03
Turkey	.27/.28	.28/.29	.86/.85	.85/.87	.76/.74	.13/.13
U.S.	.30/.38	.27/.40	.84/.87	.86/.86	.76/.77	.13/.25
M[b]	.30/.34	.30/.35	.85/.86	.85/86	.75/.77	.15/.21

Note: SENS = sensitivity, percentage of cases deviant on Child Behavior Checklist (CBCL) also deviant on Teacher's Report Form (TRF); PPV = positive predictive value, percentage of cases deviant on TRF also deviant on CBCL; SPEC = specificity, percentage of cases nondeviant on CBCL also nondeviant on TRF; NPV = negative predictive value, percentage of cases nondeviant on TRF also nondeviant on CBCL; CC = consistent classification, percentage of total cases either deviant according to both informants or nondeviant according to both informants.

[a]First entry = Internalizing. Second entry = Externalizing.

[b]Omnicultural mean for each column.

on the TRF was corroborated by a score in the deviant range on the CBCL for only about one third of the dyads having a deviant TRF score, with a range from 20% to 44% for Internalizing (positive predictive value omnicultural $M = 30\%$) and from 20% to 47% for Externalizing (positive predictive value omnicultural $M = 35\%$). In all societies studied, it was thus common for parents to not corroborate deviance based on teachers' reports and for teachers to not corroborate deviance based on parents' reports, although noncorroboration rates varied somewhat across societies.

We also computed kappa values from the 2×2 tables for Internalizing and Externalizing within each society. As shown in Table 7, kappas ranged from .03 (Thailand and Singapore) to .35 (Denmark) for Internalizing (omnicultural M kappa = .15) and from .03 (Thailand) to .35 (Jamaica) for Externalizing (omnicultural M kappa = .21). The r between the kappas for Internalizing and Externalizing across the 21 societies was .57 ($p < .007$), indicating that societies with low (or high) kappas on Internalizing also tended to have low (or high) kappas on Externalizing. After converting the 21 Internalizing and Externalizing kappas to Fisher's zs, we compared them via a paired t test. The mean kappa for Externalizing was significantly higher than that

for Internalizing, although both were low, $t(20) = 4.18$, $p < .001$.

DISCUSSION

By testing parent–teacher agreement in 21 diverse societies using five different methods, our study contributes both to theoretical understanding of cross-informant agreement and to clinical understanding of informant differences.

Problem Scale Scores

We first tested whether parents report more problems than teachers, and whether this varies by type of problem and by society. As hypothesized, scores from parents' ratings were significantly higher than scores from teachers' ratings on most problem scales, with informant ESs from <1% to 17% on the different scales. The small Informant × Society interactions (1–5%) indicate that the societies studied varied only modestly in the degree to which CBCL scores were higher than TRF scores. An important finding for both theoretical understanding and clinical practice is that the informant ES was smaller for Internalizing than for Externalizing, particularly for

younger children (4% vs. 10% for ages 6–11 and 4% vs. 6% for ages 6–16), although CBCL scores were higher than TRF scores for both kinds of problems. This indicates that ratings of internalizing kinds of problems (anxiety, sadness, somatic symptoms) were rather similar for parent and teacher reports, whereas parents tended to rate their children as having considerably more externalizing kinds of problems (aggression and rule-breaking behavior) than did teachers. Although our study cannot determine the reasons for this finding, it could reflect home–school variation in triggers and/or sanctions for externalizing behavior or differences in decision thresholds (De Los Reyes et al., 2013).

Consistent with other studies (e.g., Rescorla et al., 2013), significant gender effects (1–4%) were found on many scales, with boys scoring higher on scales tapping attention problems and externalizing problems and girls scoring higher on internalizing problems. Although many age effects were significant, no age effects or interactions of age or gender with informant exceeded 1%. This was the case both for empirically based syndromes and for *DSM*-oriented scales, despite the fact that these two sets of scales partition problems somewhat differently (e.g., anxiety and depression are combined in the Anxious/ Depressed syndrome but separated in the *DSM*-oriented scales). It is possible that age effects were smaller in this study than might be expected from epidemiological studies of diagnoses because we tested the variance accounted for in ANOVAs of dimensional scores rather than comparing the prevalence of diagnoses by age group.

Cross-Informant Correlations Between Scale Scores

We then addressed the question of whether parent–teacher rs varied across different problem scales and across the societies studied. The omnicultural mean r for Total Problems was .29, very similar to the parent–teacher r of .27 reported by Achenbach et al. (1987). Higher cross-informant rs were not found for scales comprised largely of items that were seldom endorsed by either parents or teachers (e.g., the Thought Problems scale, which includes items such as *Sees things that are not there*.) An important finding for both theoretical understanding and clinical practice is that the parent–teacher r was significantly higher for Externalizing than Internalizing (.33 vs. .24), despite the fact that there was a larger CBCL > TRF score discrepancy for the former than the latter. Our findings thus suggest that, although Externalizing problems are more visible and hence easier for teachers and parents to observe than Internalizing problems, teachers see lower levels of Externalizing problems than parents do. Among the narrow-band scales, the best agreement was for the relatively overt Attention Problems and *DSM*-Attention Deficit Hyperactivity

Problems scales, whereas the worst agreement was for the relatively covert Thought Problems scale.

The societies studied differed widely in mean cross-informant r (from .09 to .49), with the lowest rs in Thailand, Singapore, Hong Kong, Turkey, China, and Iran. When mean CBCL Total Problems scores were inspected for the nine societies with cross-informant rs $\geq$ .29, three had low mean scores (Serbia, Croatia, and Finland, 14.4–16.8, $M = 15.8$), three had moderate mean scores (Netherlands, Romania, and Denmark, 19.9–23.3, $M = 21.4$), and three had high mean scores (Lithuania, Portugal, and Puerto Rico, 25.1–32.0, $M = 28.6$) on CBCL Total Problems.

To explore possible reasons for societal differences in cross-informant rs, we were able to obtain data for 18 of the 21 societies on five survey questions (class size, plus three-level scales for degree of classroom structure, parental engagement in the school/classroom, parent–teacher phone/e-mail/mail contacts, and in-person parent–teacher interaction). Although the ratings provided on these questions are of unknown validity, some interesting patterns emerged; data are available from the first author upon request. The largest classes (approaching 40 children) were reported for Singapore, China, Thailand, Hong Kong, and Iran, which tended to have low agreement, whereas the smallest class sizes (15–25) were reported for Denmark, which had the highest agreement. Except for the Netherlands, Denmark, Finland, Croatia, and the United States, where classrooms were reported to have a "moderate" degree of structure, classrooms in the other 13 societies were described as "very structured" (e.g., China = "very structured for teacher-directed work, formal, and tight schedule and lots of homework after school," Jamaica = "highly teacher-directed work = 3," Serbia = very much structured"). Parent involvement in the classroom was reported to be highest in the Netherlands and the United States and lowest in Denmark, China, Hong Kong, Serbia, France, and Romania. The lowest level of parent–teacher contact was reported for China, Hong Kong, and Romania. Overall, it appeared that large class size was the characteristic most associated with lower parent–teacher agreement.

Agreement for Mean Item Ratings

The third question addressed how well parents and teachers in different societies agreed on item ratings. Based on mean item ratings within each society, the predicted strong parent–teacher agreement was found in every society studied regarding which items received low, medium, or high ratings (omnicultural mean Q correlation = .77), consistent with CBCL–YSR findings (Rescorla et al., 2013). As Table 6 shows, items with the highest mean ratings reflect a variety of problems. It is noteworthy that the largest societal ESs were for

Item 32. *Feels he/she has to be perfect* = 22%, Item 71. *Self-conscious or easily embarrassed* = 16%, and Item 69. *Secretive or keeps things to self* = 15%. This suggests striking cultural differences in these internalizing characteristics, a pattern consistent with the fact that the largest societal ES was found for the Anxious/Depressed syndrome (18%).

Dyadic Agreement on Item Ratings

The fourth question addressed how well parent–teacher dyads agreed on which items tended to receive low, medium, or high ratings. As predicted, the omnicultural mean for within-society mean dyadic Q was modest (.23, range = .17–.30), with minor between-society variation but great within-society variation regarding item agreement. An important implication for practice is that although clinicians are likely to encounter some parent–teacher dyads that have strong cross-informant item agreement, it is quite common for dyads to have weak item agreement. The great within-society variation in dyadic agreement suggests that degree of dyadic cross-informant agreement may have some diagnostic utility, such as indexing the degree of cross-situational consistency in a child's problems, as suggested by Dirks et al. (2012).

Cross-Informant Agreement on Deviance Status

The final question addressed how well parents and teachers in 21 societies agreed on whether the child had enough problems to exceed a deviance threshold. In all societies studied, most children who were nondeviant on the CBCL were also nondeviant on the TRF (*M* specificity = 85% for Internalizing and 86% for Externalizing). In addition, most children who were nondeviant on the TRF were also nondeviant on the CBCL (*M* negative predictive value = 85% for Internalizing and 86% for Externalizing). However, a deviant CBCL score was corroborated by a deviant TRF score for only about one third of the dyads (i.e., specificity) and a deviant TRF score was corroborated by a deviant CBCL score for only about one third of the dyads (i.e., positive predictive value), with modest variation across societies studied. Furthermore, in all societies studied, the rate of corroboration of deviance by parents was very similar to the rate of corroboration of deviance by teachers.

Societal Similarities and Differences

Our study revealed numerous similarities in parent–teacher agreement among the societies studied: CBCL scores tended to be higher than TRF scores in almost all societies studied; age and gender effects on problem scores were similar across societies studied; cross-informant agreement tended to be best for scales tapping attention and externalizing kinds of problems; there was very good parent–teacher agreement in every society studied on the items rated low, medium, or high; dyadic agreement on item ratings tended to be modest and highly variable in every society studied; and dyadic agreement on deviance status was similar across societies studied.

On the other hand, significant societal differences were also found. The magnitude of the CBCL > TRF informant effect varied across societies studied: Singapore, Portugal, Hong Kong, and Puerto Rico had large CBCL > TRF mean differences; Jamaica and Turkey had small CBCL > TRF mean differences; and Thailand showed the reverse pattern. No obvious cultural characteristic (e.g., East vs. West) appeared to explain these modest societal differences in CBCL > TRF discrepancy. Societal differences were also found in scale score cross-informant *r*s, but mean cross-informant *r*s were very similar to *r*s in the Achenbach et al. (1987) meta-analysis and in several single-society studies. Although our survey of classroom characteristics was only a first step toward understanding the cultural factors that might account for variations in parent–teacher agreement, our results suggest that larger class size may be associated with lower levels of agreement.

Limitations

Unknown variations in translation quality may have contributed error to our findings, although the large correlations on mean item ratings suggest that parents and teachers interpreted the problem items quite similarly within and across societies. Another limitation is that our study used an *etic* approach, whereby the same instruments were used in each society (Pike, 1967), rather than an *emic* approach focused on interpretations of the problem items in each society. Although our brief survey suggested that larger class size might be associated with lower levels of agreement, degree of classroom structure did not seem to be related to agreement levels, probably because 13 of 18 investigators described classrooms in their societies as very structured, including those with low (Singapore), moderate (Jamaica), and high (Serbia) levels of agreement. Additional limitations include that socioeconomic status information was not available; sample sizes, data collection methods, response rates, and inclusion of referred children varied across societies studied; and some samples were regional rather than national. Furthermore, because pre-2001 versions of the CBCL and TRF were used in many societies, six items had to be excluded. In addition, unlike De Los Reyes et al. (2009), we did not have any measures of children's behavior other than parent and teacher reports (e.g., a "criterion" measure such as behavioral observations at home and school). We could not obtain observational data for more than 27,000 children in 21 societies, but, like informant data,

observational data have limitations and error. Nonetheless, observational data would have allowed us to test with more certainty whether informant discrepancies reflected contextual differences in children's behavior. Our study also lacked outcome data for the participants. As Loeber, Green, Lahey, and Stouthamer-Loeber (1991) and Youngstrom, Findling, and Calabrese (2003) found, outcome data can reveal significant differential validity of ratings obtained from parents, teachers, and children themselves. Finally, additional research on parent–teacher agreement in clinical samples is needed to test the generalizability of our findings.

Conclusions and Implications

Our central finding—that rather modest agreement between parents and teachers on problem scores is not just a U.S. phenomenon but is quite general across the many diverse societies studied—provides international replication for meta-analytic findings first summarized by Achenbach et al. (1987) and updated by De Los Reyes and Kazdin (2005). Our results also indicate that parent–teacher agreement was higher for externalizing and attention problems than for internalizing problems. Furthermore, parent–teacher agreement regarding which items tended to receive low, medium, or high ratings was very good in every society studied. Parents and teachers converged very well on which items they rated highest, a diverse set of problems from several syndromes. Another important finding was that parent–teacher dyads within each society studied varied widely in their item agreement, which supports the suggestion by Dirks et al. (2012) that degree of cross-informant agreement may be a good index of the pervasiveness of behavioral symptoms across contexts. Finally, parent noncorroboration of deviance rated by teachers and teacher noncorroboration of deviance rated by parents were common in the societies studied and comparable in magnitude (omnicultural mean of about 70% for each direction). This finding suggests that both the "and" rule and the "or" rule may obscure important clinical information about the contextual factors associated with the modest agreement between teachers and parents regarding children's behavior.

Clinicians who work with adolescents will find it useful to compare the findings of the current study with Rescorla et al.'s (2013) findings for CBCL–YSR agreement in 25 societies. In both studies, informant differences were significant but varied in magnitude across societies studied. However, the mean YSR Total Problems score was higher than the mean CBCL Total Problems score in every society studied, whereas the mean CBCL Total Problems score was higher than the mean TRF Total Problems score in all societies studied but Thailand. Also, informant ESs varied considerably across different kinds of problems in both studies. CBCL–YSR informant ESs were large and similar in size for Internalizing and Externalizing (16% vs. 18%), whereas CBCL-TRF informant ESs for these scales were smaller for both age groups (ages 6–11: 4% vs. 10%; ages 6–16: 4% vs. 6%) and dissimilar in size for ages 6 to 11. Mean cross-informant rs for problem scale scores averaged .41 for the CBCL–YSR but .29 for the CBCL–TRF, indicating better parent–adolescent agreement than parent–teacher agreement. In addition, in the YSR study, mean cross-informant rs were similar for Internalizing (.45) and Externalizing (.46), whereas in the current study parent–teacher agreement was significantly greater for Externalizing than Internalizing (.33 vs. .24). In both studies, Q correlations between mean item ratings were large in every society studied, indicating that parents, on average, agreed well with both adolescents and teachers with respect to which problem items they rated low, medium, or high. Also consistent across both studies was the great variation in dyadic agreement and the low mean dyadic agreement level within every society studied. Moreover, in both studies, noncorroboration of deviance between informants was common and symmetrical in every society. However, both parental noncorroboration of self-reported deviance and adolescent noncorroboration of parent-reported deviance averaged about 50%, whereas both parental noncorroboration of teacher-reported deviance and teacher noncorroboration of parent-reported deviance averaged about 70%. Whether such high parent–teacher noncorroboration rates would be found for clinical samples is an important topic for future research.

De Los Reyes et al. (2012) and Achenbach (2011) have argued that modest cross-informant agreement may not just reflect measurement error but may indicate variability in children's behavior across contexts, informants' differential knowledge of children's behaviors, informant differences in "decision" thresholds, and transactional effects between informants and children that elicit different behaviors. These sources of variance underscore the importance of routinely obtaining data from both parents and teachers and preserving the information obtained so that informant reports can be systematically compared. Equally important, our findings underscore the importance of measuring cross-informant agreement in multiple ways. Consistent with the arguments of Dirks et al. (2012), our findings suggest that a high level of informant agreement suggests cross-situational consistency of problems.

Although we did not directly test contextual factors that might lead to low levels of parent–teacher agreement, we can suggest some potentially relevant factors. For example, when TRF scores are much higher than CBCL scores, clinicians should consider whether academic or social issues contribute to the child's difficulties at school. Conversely,

when CBCL scores are much higher than TRF scores, clinicians should consider whether family dynamics contribute to the child's difficulties at home. Such clinical hypotheses can be explored by obtaining behavioral observations in the classroom, observations of parent–child interactions during clinical interviews, and interviews with teachers about classroom issues. As De Los Reyes et al. (2009) demonstrated, such additional data can provide important clarification of parent–teacher discrepancies in clinical cases.

REFERENCES

Achenbach, T. M. (2011). Commentary: Definitely more than measurement error: But how should we understand and deal with informant discrepancies? *Journal of Clinical Child and Adolescent Psychology*, *40*, 80–86. doi:10.1080/15374416. 2011.533416

Achenbach, T. M., Bird, H. R., Canino, G. J., Phares, V., Gould, M., & Rubio-Stipec, M. (1990). Epidemiological comparisons of Puerto Rican and U.S. mainland children: Parent, teacher, and self reports. *Journal of the American Academy of Child and Adolescent Psychiatry*, *29*, 84–93. doi:10.1097/00004583-199001000-00014

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *101*, 213–232. doi:10.1037//0033-2909.101.2.213

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. Burlington: University of Vermont, Research Center for Children, Youth, and Families.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Bilenberg, N. (1999). The Child Behavior Checklist (CBCL) and related material: Standardization and validation in Danish population and clinically based samples. *Acta Psychiatrica Scandivica, Supplementum*, *100*, 398, 1–52. doi:10.1111/j.1600-0447.1999.tb10703.x

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Academic Press.

De Los Reyes, A. (2011). Introduction to the special section: More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, *40*, 1–9. doi:10.1080/153744 16.2011.533405

De Los Reyes, A., Henry, D. B., Tolan, P. H., & Wakschlag, L. S. (2009). Linking informant discrepancies to observed variations in young children's disruptive behavior. *Journal of Abnormal Child Psychology*, *37*, 637–652. doi:10.1007/s10802-009-9307-3

De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, *131*, 483–509. doi:10.1037/0033-2909.131.4.483

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

Deng, S., Liu, X., & Roosa, M. W. (2004). Agreement between parent and teacher reports on behavioral problems among Chinese children. *Journal of Developmental and Behavioral Pediatrics*, *25*, 407–414. doi:10.1097/00004703-200412000-00004

Dirks, M. A., De Los Reyes, A., Briggs-Gowan, M., Cella, D., & Wakschlag, L. S. (2012). Annual research review: Embracing not erasing contextual variability in children's behavior–theory and utility in the selection and use of methods and informants in developmental psychopathology. *Journal of Child Psychology and Psychiatry*, *53*, 558–574. doi:10.1111/j.1469 -7610.2012.02537.x

Erol, N., & Simsek, Z. (1997). Mental health profiles of Turkey: Behavioral and emotional problems and competencies in Turkish children (in Turkish). In N. Erol, C. Kilic, M. Ulusoy, M. Kececi, & Z. Simsek (Eds.), *Mental health profiles in Turkey: A preliminary report* (pp. 12–33). Ankara, Turkey: Aydogdu.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavioral Research Methods*, *41*, 1149–1160. doi:10.3758/BRM.41.4.1149

Fombonne, E., & Vermeersch, S. (1997). Les enfants de la Cohorte GAZEL: I – Prévalence des contacts avec le système médico-éducatif pour des motifs psychologiques, et facteurs associés. *Revue Epidémiologique et Santé Publique*, *45*, 29–40.

Fonseca, A. C., Simoes, A., Rebelo, J. A., Ferreira, J. A., Cardoso, F., & Temudo, P. (1995). Hyperactivity and conduct disorder among Portuguese children and adolescents: Data from parents' and teachers' reports. In J. Sergeant (Ed.), *Eunethydis: European approaches to hyperkinetic disorder* (pp. 115–129). Zürich, Switzerland: Fotorotar.

Frigerio, A., Cattaneo, C., Cataldo, M., Schiatti, A., Molteni, M., & Battaglia, M. (2004). Behavioral and emotional problems among Italian children and adolescents aged 4 to 18 years as reported by parents and teachers. *European Journal of Psychological Assessment*, *20*, 124–133. doi:10.1027/1015-5759.20.2.124

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., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *The British Journal of Psychiatry*, *177*, 534–539. doi:10.1192/bjp.177.6.534

Hanley, J. A., Negassa, A., Edwardes, M. D., & Forrester, J. E. (2003). Statistical analysis of correlated data using generating estimating equations: an orientation. *American Journal of Epidemiology*, *157*, 364–375. doi:10.1093/aje/kwf215

Kirk, R. R. (1995). *Experimental design procedures for the behavioral sciences*. Pacific Grove, CA: Brookside, Cole.

Kraemer, H. C. (1992). *Evaluating medical tests: Objective and quantitative guidelines*. Newbury Park, CA: Sage.

Lambert, M. C., Lyubansky, M., & Achenbach, T. M. (1998). Behavioral and emotional problems among adolescents of Jamaica and the United States: Parent, teacher, and self-reports for ages 12 to 18. *Journal of Emotional and Behavioral Disorders*, *6*, 180–187. doi:10.1177/106342669800 600306

Liu, X. C., Sun, Z., Neiderhiser, J., Uchiyama, M., Okawa, M., & Rogan, W. (2001). Behavioral and emotional problems in Chinese adolescents: Parent and teacher reports. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*, 828–836. doi:10.1097/00004583-200107000-00018

Loeber, R., Green, S. M., Lahey, B. B., & Stouthamer-Loeber, M. (1991). Differences and similarities between children, mothers, and teachers as informants on disruptive child behavior. *Journal of Abnormal Child Psychology*, *19*, 75–95. doi:10.1007/BF00910566

Phillipson, S., & Phillipson, S. N. (2007). Academic expectations, belief of ability, and involvement by parents as predictors of child achievement: A cross-cultural comparison. *Educational Psychology*, *27*, 329–348. doi:10.1080/01443410601104130

Pike, K. L. (Ed.). (1967). *Language in relation to a unified theory of structure of human behavior* (2nd ed.). The Hague, Netherlands: Mouton.

Pong, S., & Pallas, A. (2001). Class size and eighth-grade math achievement in the United States and abroad. *Educational Evaluation and Policy Analysis*, *23*, 251–273. doi:10.3102/0162373 7023003251

Rescorla, L., Ginzburg, S., Achenbach, T. M., Ivanova, M. Y., Almquist, F., Begovac, I., . . . Verhulst, F. C. (2013). Cross-informant agreement between parent-reported and adolescent self-reported problems in 25 societies. *Journal of Clinical Child and Adolescent Psychology*, *42*, 262–273. doi:10.1080/15374416.2012.717870

Roussos, A., Karantanos, G., Richardson, C., Hartman, C., Karajiannis, D., Kyprianos, S., . . . Zoubou, V. (1999). Achenbach's Child Behavioral Checklist and Teacher's Report Form in a normative sample of Greek children 6–12 years old. *European Child and Adolescent Psychiatry*, *8*, 165–172. doi:10.1007/s007870050125

Rudan, V., Begovac, I., Szirovicza, L., Filipovic, O., & Skocic M. (2005). The Child Behavioral Checklist, Teacher's Report Form and Youth Self Report form problem scales in a normative sample of Croatian children and adolescents aged 7–18. *Collegium Antropologicum*, *29*, 17–26.

Stephenson, W. (1935). Correlating persons instead of tests. *Character & Personality*, *4*, 17–24. doi:10.1111/j.1467-6494.1935.tb02022.x

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

van der Ende, J., Verhulst, F. C., & Tiemeier, H. (2012). Agreement of informants on emotional and behavioral problems from childhood to adulthood. *Psychological Assessment*, *24*, 293–300. doi:10.1037/a0025500

Weisz, J. R., Suwanlert, S., Chaiyasit, W., Weiss, B., Achenbach, T. M., & Trevathan, D. (1987). Epidemiology of behavioral and emotional problems among Thai and American children: Parent reports for ages 6–11. *Journal of the American Academy of Child and Adolescent Psychiatry*, *26*, 890–897. doi:10.1097/00004583-198726060-00014

Weisz, J. R., Suwanlert, S., Chaiyasit, W., Weiss, B., Weiss, Achenbach, T. M., & Trevathan, D. (1989). Epidemiology of behavioral and emotional problems among Thai and American children: Teacher reports for ages 6–11. *Journal of the American Academy of Child and Adolescent Psychiatry*, *30*, 471–484. doi:10.1111/j.1469-7610.1989.tb00260.x

Woo, B. S. C., Ng, T. P., Fung, D. S. S., Chan, Y. H., Lee, Y. P., Koh, J. B., & Cai, Y. (2007). Emotional and behavioural problems in Singaporean children based on parent, teacher and child reports. *Singapore Medical Journal*, *48*, 1100–1106.

Wright, J. C., & Zakriski, A. L. (2001). A contextual analysis of externalizing and mixed syndrome boys: When syndromal similarity obscures functional dissimilarity. *Journal of Consulting and Clinical Psychology*, *69*, 457–470.

Youngstrom, E. A., Findling, R. L., & Calabrese, J. R. (2003). Who are the comorbid adolescents? Agreement between psychiatric diagnosis, parent, teacher, and youth report. *Journal of Abnormal Child Psychology*, *31*, 231–245.