J Psychopathol Behav Assess (2016) 38:381–397
DOI 10.1007/s10862-016-9541-2

Collateral Reports and Cross-Informant Agreement about Adult Psychopathology in 14 Societies

Leslie A. Rescorla[1] · Thomas M. Achenbach[2] · Masha Y. Ivanova[2] · Lori V. Turner[2] · Hervör Árnadóttir[3] · Alma Au[4] · J. C. Caldas[5] · Yi-Chuen Chen[6] · Jeroen Decoster[7] · Johnny Fontaine[7] · Yasuko Funabiki[8] · Halldór S. Guðmundsson[3] · Patrick Leung[9] · Jianghong Liu[10] · Jelena Srdanović Maraš[11] · Jasminka Marković[11] · Kyung Ja Oh[12] · Marina M. da Rocha[13] · Virginia C. Samaniego[14] · Edwiges Silvares[15] · Roma Simulioniene[16] · Elvisa Sokoli[17] · Natalia Vazquez[14] · Ewa Zasepa[18]

Published online: 10 March 2016

Abstract To advance international mental health assessment, instruments that have been internationally validated are needed. To this end, we analyzed ratings from 14 societies on the Adult Behavior Checklist (ABCL), a collateral-report form parallel to the Adult Self-Report (ASR; Achenbach and Rescorla 2003) for ages 18 to 59. Both the ABCL and the ASR assess problems, personal strengths, and adaptive functioning. For a sample of 8322 see note below collaterals, we found strong consistency across societies regarding which ABCL problem items tended to obtain relatively low, medium, or high ratings. Most societal effect sizes (ESs) for problem scale scores were small to medium (< 13.9 %), but the ES for the ABCL Personal Strengths scale was 25 %. For most of the same participants (N = 8,302), we analyzed cross-informant agreement between self-reports on the ASR and collateral reports on the ABCL. Cross-informant correlations for problem scale scores averaged .47, with considerable societal variation. Problem score means were higher on the ASR than the ABCL in every society, but the size of the difference varied across societies. Mean item ratings on the ABCL and ASR were highly correlated within every society (mean r = .92), but within-dyad item rating agreement varied widely in every society (mean r = .39). In all societies, non-corroboration of self-reported deviance and of collateral-reported deviance was common. Overall findings indicated considerable similarity but also some important differences in collateral-reported problems and adaptive functioning across 14 societies.

Keywords Adult psychopathology · ABCL · ASR · Collateral-reported problems · Cross-cultural · International cross-informant agreement

✉ Leslie A. Rescorla
lrescorl@brynmawr.edu

[1] Bryn Mawr College, 101 N. Merion Avenue, Bryn Mawr, PA 19010, USA

[2] University of Vermont, Burlington, VT, USA

[3] University of Iceland, Reykjavík, Iceland

[4] Hong Kong Polytechnic University, Hung Hom, Hong Kong

[5] Instituto Superior de Ciencias da Saude-Norte, Gandra, Portugal

[6] National Chung-Cheng University, Chiayi County, Taiwan

[7] Ghent University, Ghent, Belgium

[8] Kyoto University, Kyoto, Japan

[9] The Chinese University of Hong Kong, Shatin, Hong Kong

[10] University of Pennsylvania, Philadelphia, PA, USA

[11] Clinic for Psychiatry, Clinical Center of Vojvodina, Novi Sad, Serbia

[12] Yonsei University, Seoul, South Korea

[13] Mackenzie Presbyterian University, São Paulo, Brazil

[14] Pontifica Universidad Católica Argentina, Buenos Aires, Argentina

[15] University of Sao Paulo, São Paulo, Brazil

[16] Klaipeda University, Klaipėda, Lithuania

[17] University of Tirana, Tirana, Albania

[18] The Maria Grzegorzewska Academy of Special Education, Warsaw, Poland

Informant Discrepancies

Multi-informant assessment has become standard practice in child clinical services. This is because a large body of research has documented important discrepancies between reports by different informants about children (Achenbach et al. 1987; De Los Reyes 2011; De Los Reyes et al. 2015; De Los Reyes and Kazdin 2005; Rescorla et al. 2013). As noted by Dirks et al. (2012), the modest level of cross-informant agreement between different kinds of informants regarding children's behavior presents a challenge for the assumption that assessment methods such as rating scales, interviews, and observational instruments "emphasize psychopathology as a trait that will generalize across situations" (p. 560). However, De Los Reyes et al. (2013b) argued that, rather than merely reflecting measurement error, modest cross-informant agreement reveals important variability in children's behavior across contexts, such as home versus school.

Perhaps because clinicians assume that adult clients know best what problems they are having, multi-informant assessment is much rarer in adult clinical services than in children's services. Furthermore, as noted by De Los Reyes et al. (2013a), research on cross-informant agreement regarding adult mental health is much less common than research regarding child mental health. In their study of adult cross-informant agreement, De Los Reyes et al. found better client–clinician agreement for adults with generalized than with non-generalized social anxiety disorders. Better agreement was also found for adults who exhibited social skills deficits that were consistent across three social interaction tasks than inconsistent across the tasks.

One of the few systematic reviews of cross-informant agreement in adults is a meta-analysis by Achenbach et al. (2005). Achenbach et al. reported that their search for correlations between adult self-reports and collateral reports in 51,000 articles published in 52 peer-reviewed journals yielded only 108 studies (0.2 % of the articles searched), indicative of the paucity of research in this area. Their meta-analysis demonstrated important cross-informant disparities regarding adult psychopathology, parallel to the patterns Achenbach et al. (1987) reported for children decades earlier. Specifically, Achenbach et al. (2005) reported mean correlations (*r*s) between collaterals' ratings and adults' self-ratings of only .43 for internalizing problems (e.g., anxiety, depression) and .44 for externalizing problems (e.g., aggression, lying). Moreover, Meyer et al. (2001) reported that 70 % of psychiatric diagnoses of adults based on self-reports were inconsistent with diagnoses made using collateral reports. *Kappas* between diagnoses made from self-reports versus multiple sources of data ranged from only .12 to .34 (mean = .18).

There are numerous reasons why collateral reports differ from self-reports. As suggested by De Los Reyes et al. (2013a,b), modest cross-informant agreement could reflect contextual variation in an individual's behavior, such as at home versus at work. Additionally, discrepancies between collateral reports and self-reports could reflect differences in attributions, such as collateral reports reflecting personological/trait attributions but self-reports reflecting situational attributions (Ross 1977). Furthermore, differences between collateral reports and self-reports could reflect the fact that collaterals cannot observe the adult's behavior in situations in which they are not present. Adults may also be unaware of how others view their behavior (e.g., a father thinks his adult son is being argumentative, while the son thinks he is just joking). Finally, adults have various collaterals who could report on their functioning, such as a spouse/partner, parent, sibling, adult child, friend, or colleague. Cross-informant agreement might differ with the degree of intimacy in the adult-collateral relationship. For example, a spouse/partner might agree better with a target adult than that adult's parent, child, sibling, or friend.

Standardized Adult Assessment

To facilitate multi-informant assessment of adults, it is necessary to have instruments for obtaining parallel self-reports and collateral reports. For this purpose, Achenbach and Rescorla (2003) developed the Adult Behavior Checklist (ABCL) as a collateral-report form paralleling the Adult Self-Report (ASR; Achenbach and Rescorla 2003). Both are low-cost, standardized rating forms written at a fifth grade reading level to assess behavioral, emotional, social, and thought problems and adaptive functioning for adults ages 18–59. The ABCL obtains ratings from collaterals for 118 problem items, 115 of which are also on the ASR. For both the ABCL and the ASR, respondents rate problem items as *0 = not true*, *1 = somewhat or sometimes true*, and *2 = very true or often true* based on the preceding 6 months.

As detailed by Achenbach and Rescorla (2015), ABCL and ASR item ratings yield scores on eight statistically derived narrow-band syndromes (e.g., Aggressive Behavior); three broad-band scales (e.g., Internalizing Externalizing, and Total Problems); six *DSM*-oriented scales (e.g., Antisocial Personality Problems); and two scales based on research conducted by others (Obsessive-Compulsive Problems and Sluggish Cognitive Tempo). The ABCL and ASR are also scored on a Substance Use scale, but substance use findings are not presented in this report. Interspersed among the problem items are 11 Personal Strengths items (e.g., *I make good use of my opportunities*), which are rated on the same 0–1-2 scale as the 118 problem items. Both the ABCL and the ASR have Friends and Spouse/Partner scales, while the ASR also has Family, Job, and Education scales not included on the ABCL.

The feasibility of collecting collateral information about adults was demonstrated in the national household survey that produced the U.S. ASR-ABCL norms, wherein 2020 18–59-year-olds who completed the ASR were asked to nominate a collateral to complete the ABCL (Achenbach and Rescorla 2003). ABCLs were obtained for 1636 ASR cases (81.0 % of the ASR sample). Although adults received $10 for completing an ASR or an ABCL, it seems likely that high completion rates would also be found in clinical and research settings if adult clients asked a family member or friend to fill out an ABCL.

International Epidemiology

Mental disorders comprise about 14 % of the global health burden worldwide, are linked to many other health problems, and are among the most costly disorders to treat (Tomlinson et al. 2009). As Tomlinson et al. noted, most mental health research has been done in high-income Western societies, which raises questions about the generalizability of those findings to other societies. To test whether findings in one society are generalizable to other societies, the same assessment instrument must be used in different societies, employing what Pike (1967) called the *etic* approach to research. Etic research can reveal similarities and differences between societies in the prevalence of categorically defined psychiatric disorders (such as depression) and in quantitative scores on rating scales (such as scales measuring depression).

To our knowledge, there are no multinational comparisons of adult psychiatric diagnoses based on collateral reports, comparable to the World Health Organization (WHO 2004) multinational study of diagnoses based on self-reports obtained in interviews. In that study, diagnoses of anxiety disorders, mood disorders, impulse-control disorders, and substance disorders were made according to *DSM-IV* criteria (American Psychiatric Association 1994). Completion rates ranged from 46 % (France) to 88 % (Colombia). Prevalence estimates for ≥1 diagnosis ranged from 4 % in Shanghai to 26 % in the U.S.

A more cost-effective method than diagnostic interviews for international epidemiological studies is to use standardized rating instruments such as the ASR and ABCL. Ivanova et al. (2015b) used confirmatory factor analyses (CFAs) to test the generalizability of the U.S.-derived 8-syndrome model to ASR self-ratings by 18- to 59-year-olds in 29 societies (N = 17,152). The primary fit index (Root Mean Square Error of Approximation, RMSEA) indicated good fit between the data and the 8-syndrome model for all 29 samples. In a companion study, Rescorla et al. (2016) analyzed self-ratings of problems and adaptive functioning on the ASR for population samples of adults from 17 of the Ivanova et al. societies plus the U.S. (N = 12,217). Results indicated strong consistency across societies for problem scale alphas and for the ASR problem items that obtained low, medium, or high ratings. Ten of the 18 societies had mean Total Problems scores within one *SD* (6.0) of the omnicultural mean of 42.5. Note that we use the term "omnicultural mean" as defined by Ellis and Kimmel (1992) to refer to the average of the means for the 18 societies studied. Effects of society for most problem scales were small (2–5 %), but the societal effect on the ASR Personal Strengths scale was very large (34 %). Scores indicative of poor adaptive functioning with spouse/partner and in job and educational settings were significantly associated with high problem scores.

To our knowledge, the only published international comparison of collateral reports for adults is Ivanova et al.'s (2015a) CFA study testing the ABCL's 8-syndrome model in informants' ratings for 8582 adults ages 18 to 59 in 18 societies. In the current study, we used 14 of the 18 ABCL data sets Ivanova et al. analyzed, plus the U.S. normative sample, from which the factor model Ivanova et al. tested was derived. The four Ivanova et al. data sets we excluded (from Algeria, France, Russia, and the United Kingdom) were not sufficiently representative of the 18–59 age range to be used for normative comparisons (e.g., they included only university students or some other narrow age group).

Ivanova et al. (2015a) reported that the RMSEA, which was the primary model fit index, showed good fit for all societies, while secondary indices (Tucker Lewis Index, Comparative Fit Index) showed acceptable to good fit for 17 societies. Factor loadings were robust across societies and items. Ivanova et al. (2015a) reported that only 4 (0.08 %) of the 5007 parameters were outside the admissible parameter space. Median factor loadings ranged from .53 to .79, with an overall median of .70. The findings were consistent with CFA findings reported by Ivanova et al. (2015b) for the ASR, which were based on self-ratings in 29 societies. Fourteen of the ASR data sets Ivanova et al. (2015b) analyzed were also used in the current study to analyze cross-informant agreement, but the other 15 data sets were excluded because the samples were not representative of the full 18–59 year age range or did not include ABCLs for the same individuals as the ASRs. The results of Ivanova et al.'s two studies thus support the generalizability of the ASR-ABCL 8-syndrome model for characterizing adult psychopathology in diverse societies and are directly relevant for the present study because they analyzed the same data sets we analyzed.

Rationale and Goals for our Study

Collateral report instruments such as the ABCL can broaden assessment of adults' strengths and problems by augmenting self-reports with reports by people who know the person being

assessed. However, to our knowledge, there are very few published studies of collateral reports about adults' mental health problems. Our study examining collateral reports by 8322 adults in 14 societies therefore addresses a significant limitation in international mental health research. Furthermore, discrepancies between self- and collateral ratings (Achenbach et al. 2005) and between diagnoses made from self-reports versus collateral reports (Meyer et al. 2001) indicate that collaterals often provide data not obtained from self-reports. However, as indicated by the Achenbach et al. (2005) meta-analysis, research on cross-informant agreement for adults is very limited, and no international comparisons of agreement between self-reports and collateral reports have been published. Our study of adult cross-informant agreement in 14 societies is thus an important advance.

The first major goal of our study was to identify similarities and differences across societies in collateral reports of adults' problems, personal strengths, and adaptive functioning on the ABCL. The second major goal was to test cross-informant agreement between collateral reports on the ABCL and self-reports on the ASR. Our specific research questions addressing these major goals are detailed below.

ABCL analyses For our ABCL analyses, we used data from 14 societies differing widely in economic, political, ethnic, religious, and cultural characteristics (N = 8322) to answer two central questions: (a) How similar are the 14 societies with respect to which ABCL problem items tend to obtain low, medium, or high ratings? and (b) What are the effects of society, gender, age, and rater type (spouse/partner vs. others) on ABCL scale scores? Prior to addressing the second question, we examined the internal consistencies of the ABCL scales across societies.

ABCL-ASR cross-informant analyses For our cross-informant analyses, we compared ABCL and ASR data from 14 societies (N = 8,203) to answer the following questions regarding cross-informant agreement: (a) How similar are the 14 societies with respect to correlations between ASR and ABCL scale scores, and do these correlations differ by the type of rater who completed the ABCL? (b) Do ABCL scores differ significantly from ASR scores, and do informant effects vary by society, gender, age, rater type, and problem type? (c) How similar are the 14 societies with respect to agreement between target participants and collaterals, on average, regarding which items obtain low, medium, and high ratings? (d) What is the variation within and between societies regarding how well dyads agree on item ratings, and does this differ by rater type? (e) How similar are the 14 societies with respect to dichotomous agreement between collateral ratings and self-ratings regarding deviance status?

Method

Samples

Table 1 contains information about the 15 samples we analyzed. As explained below, we used 14 samples for international comparisons of ABCL scores (not China) and 14 samples for the ABCL-ASR cross-informant analyses (not Iceland). Fourteen of the 15 samples were among the 18 ABCL samples Ivanova et al. (2015a) utilized for CFAs, with the 15th being the U.S, for a total sample of 8322 for our ABCL analyses. The four samples we omitted were not sufficiently representative of the 18–59 age range to be used for normative comparisons of ABCL scale scores and item ratings (e.g., they were all university students or all parents of 9-year-olds, etc.). As shown in Table 1, rigorous random sampling methods were used in some societies, resulting in nationally representative population samples. However, in other societies, various methods of convenience sampling were used, resulting in regional samples or samples of unknown representativeness.

Because one of our 14 ABCL samples (from Iceland) contained few collateral reports of the same individuals whose self-reports were obtained on the ASR, this sample was excluded from our cross-informant analyses. However, a sample from China, although not sufficiently representative of the 18–59 age range to be used for normative comparisons, did have matched ABCL and ASR data. We therefore included the Chinese sample in our cross-informant analyses, yielding a total N of 8,203 adults from 14 societies. ABCL participants were recruited on the basis of nominations by ASR participants (except in Iceland). Collaterals included spouses/partners, family members, and friends. As shown in Table 1, the percentage of ABCL informants who were spouses or partners of the ASR adult ranged from 9 % in Hong Kong to 100 % in Korea. Because other kinds of informants were not consistently delineated in the data sets, we grouped them together as *other*. The data for each society were collected by indigenous investigators and then sent to the lead authors for analysis.

Measure

The ABCL, which takes 15–20 min to complete, includes four Friends items and seven Spouse/Partner items (completed if the respondent has lived with a spouse/partner in the past 6 months). For the Friends scale, collaterals rate the person's number of friends, frequency of contacts with friends, getting along with friends, and visits by friends and family. For the Spouse/Partner scale, items include how well the person gets along with his/her

Table 1 Source, ABCL N (cross-informant N), percent male, spouse/partner rater percent, and sampling procedure

Society	Reference	*N*	% Male	% Spouse/ Partner[c]	Sampling procedure
Albania	Sokoli (2013)[a]	750 (747)	50 %	39 %	Collaterals selected by ASR participants recruited in a representative national sample stratified by age, gender, region and urban/rural.
Argentina	Samaniego and Vázquez (2012)	679 (679)	48 %	28 %	Collaterals selected by ASR participants recruited in a regional sample stratified by level of educational attainment to be representative of the greater Buenos Aires area.
Brazil	Silvares & Rocha (2012) [a]	679 (558)	40 %	29 %	Collaterals selected by ASR participants recruited in a convenience national sample stratified by region, age, gender, and SES to be representative of the Brazilian metropolitan population.
China	Liu (2012) [a]	—[b] (454)	39 %		Collaterals selected by ASR participants recruited in a regional sample in Jintan, China.
Flanders (Belgium)	Decoster & Fontaine (2012) [a]	737 (737)	50 %	55 %	Collaterals selected by ASR participants recruited in a regional convenience sample stratified by region, gender, age, and education, to be representative of the population of Flanders (Belgium).
Hong Kong	Au & Leung (2012) [a]	330 (234)	40 %	9 %	Collaterals selected by ASR participants recruited in a community convenience sample stratified by age and gender to be representative of the Hong Kong population.
Iceland	Guðmundsson & Árnadóttir (2012) [a]	399 (NA)	46 %	67 %	Most ABCLs were completed by ASR participants for people they knew, with only a small portion being collaterals selected by ASR participants; ASR participants were recruited by national stratified random sampling to be representative of the Icelandic population.
Japan	Funabiki (2012) [a]	1,000 (1000)	47 %	85 %	Collaterals selected by ASR participants who were recruited as a Japanese national sample stratified by age and gender by a survey company.
Korea	Oh and Kim (2010)	349 (299)	50 %	100 %	Collaterals selected by ASR participants recruited in a representative ample randomly drawn from a national registry and stratified by age, gender, and educational attainment to be representative of the South Korean population.
Lithuania	Šimulionienė et al. (2010)	573 (573)	48 %	46 %	Collaterals selected by ASR participants recruited by stratified random sampling from the Lithuanian national registry, with stratification by gender, age, and education.
Poland	Zasepa (2012) [a]	281(281)	39 %	54 %	Collaterals selected by ASR participants recruited in a national sample by random sampling and stratified by age, gender, residence, and educational attainment to be representative of the Polish population.
Portugal	Caldas (2012) [a]	397 (395)	49 %	41 %	Collaterals selected by ASR participants recruited in a regional convenience sample stratified by age and gender to be representative of the Portuguese population.
Serbia	Markovic (2012) [a]	312 (312)	43 %	64 %	Collaterals selected by ASR participants recruited in a randomly drawn representative sample of the Novi Sad metropolitan area stratified by age.
Taiwan	Chen (2012) [a]	300 (300)	50 %	45 %	Collaterals selected by ASR participants in a convenience sample stratified by region, gender, and age to be representative of the Taiwan population.
U.S.	Achenbach and Rescorla (2003)	1636 (1634)	40 %	43 %	Collaterals selected by ASR participants recruited by random household sampling stratified by age, gender, ethnicity, and urban/rural to be representative of the U.S. population.

[a] Unpublished raw data

[b] Data from China were used only in the ABCL-ASR cross-informant analyses

[c] Percent of ABCL raters who were a spouse/partner of the person they rated

partner, shares responsibilities, enjoys similar activities, is satisfied with his/her partner, and likes his/her partner's friends and family. The 118 problem items and 11 interspersed Personal Strengths items are rated *0 = not true* (*as far as you know*), *1 = somewhat or sometimes true*, and *2 = very true or often true*. Foreign language versions of the ABCL (Achenbach and Rescorla 2003) were developed by indigenous mental health researchers who first did translations and then obtained independent back-translations.

Based on U.S. data, Achenbach and Rescorla (2003) reported alphas of .92 to .97 for the ABCL broad-band scales, .70 to .91 for the syndromes, and .70 to .88 for the *DSM*-oriented scales. As reported by Achenbach and Rescorla (2003), the ABCL's one-week test retest correlations (*r*s) were .80 to .92 for the broad-band scales, .73 to .88 for the syndromes, and .75 to .89 for the *DSM*-oriented scales. For U.S. data, cross-informant ABCL-ASR correlations averaged .40 for the empirically based problem scales, .38 for the *DSM*-oriented problem scales, and .48 and .54 for the Friends and Spouse/Partner scales, respectively. Achenbach and Rescorla (2003) also reported that ABCL items and scales significantly discriminated between referred versus non-referred samples of adults, thus supporting their validity.

Data Analysis

ABCL analyses ABCL problem scale scores were positively skewed in every sample, because many people in population samples have relatively few problems. However, general linear models are very robust with respect to deviations from normality, especially with large samples having similar skew and very stringent criteria for significance, such as the $p < .001$ that we used (Kirk 1995). Accordingly, we analyzed untransformed raw scale scores.

In a preliminary analysis to see if type of rater (spouse/partner vs. other) affected ABCL scores and was therefore important to include as a variable in our other ABCL analyses, we analyzed scores from the three broad-band problem scales (Total Problems, Internalizing, Externalizing), the Personal Strengths scale, and the Spouse/Partner scale using rater-type x society ANOVAs. We excluded Hong Kong and Korea from these analyses, as they had too great an imbalance in informant percentages for effects of type of rater to be tested. These ANOVAs indicated that the effect of rater type was not significant for Total Problems and Externalizing and significant at $p < .001$ but with ESs ≤1 % for the Internalizing, Spouse/Partner, and Personal Strengths scales. The society x rater-type interactions were non-significant (Internalizing, Spouse/Partner, and Personal Strengths) or significant and very small (ES <1 % for Total Problems and Externalizing). As these results indicated negligible effects of rater type for the three broad-band problems scales, as well as for the Spouse/Partner and Personal Strengths scales, we did not include rater type in our remaining ABCL analyses.

Our next analyses addressed cross-societal similarities between items receiving low, medium, or high ratings. To this end, we computed Pearson correlations between the mean of the 0–1-2 ratings for each item between each pair of societies. We refer to these correlations as *r*s but technically they are *Q* correlations (Stephenson 1935), which are Pearson *r*s computed over items (i.e., mean of the 0–1-2 ratings obtained for each item) between samples (i.e., pairs of societies). We then averaged each society's 13 bi-society *r*s, yielding a mean *r* for each society. These 13 *r*s were then averaged to yield an omnicultural mean *r* for item ratings. To identify the problem items with the highest mean ratings and the lowest mean ratings across all 14 societies, we also averaged the 118 mean problem item ratings across the 14 societies to obtain omnicultural mean item ratings.

Prior to conducting analyses of scale scores, we calculated within-society internal consistency reliability (alpha coefficients) for all 22 scales (the 19 problem scales, the Personal Strengths scale, and the Friends and Spouse/Partner scales). Pair-wise correlations between these 22 societal alphas showed that patterns of internal consistencies were quite consistent across societies. We then tested effects of society, rater type, gender, and age (18–35, 36–59) on scores for our 22 scales using analyses of variance (ANOVAs). Effect sizes (ESs) for ANOVAs were measured by η^2 and characterized using Cohen's (1988) criteria (*small* = .01 to .059, *medium* = .06 to .139, *large* ≥ .14).

ABCL-ASR Cross-Informant Analyses To achieve the second major goal of our study, namely testing cross-informant agreement between collateral reports on the ABCL and self-reports on the ASR, we conducted several additional analyses. As noted above, our cross-informant analyses involved 14 societies, yielding a total *N* of 8,203 dyads (Iceland was excluded but China was added for these analyses). These analyses were all based on the 115 problem items in common between the ASR and the ABCL.

We calculated Pearson *r*s to assess ABCL-ASR concordance for scale scores within each society, first using the full sample and then comparing *r*s by rater type (spouse/partner vs. other). We then used mixed-model ANOVAs to test scale scores for effects of society, informant, gender, and age as factors. We redid these analyses with rater type as an additional factor for the 12 societies with sufficient spouse/partner versus other types of raters to be analyzed. To limit the number of these rater-type ANOVAs, we conducted them for Total Problems, Internalizing, Externalizing, Personal Strengths scale, and Spouse/Partner only. Next, we calculated Pearson *r*s between ABCL and ASR mean item ratings, which are in effect *Q* correlations. Next, we calculated within-dyad *Q* correlations for item ratings (using Pearson *r*s) and used ANOVAs to test the effects of society, rater type, gender,

and age on these *Q*s transformed to Fisher's *z*s. Lastly, we cross-tabulated deviance status between the ASR and the ABCL and then computed decision statistics. Because of our high statistical power, we set alpha at $p \leq .001$ and report ESs rather than *F* and *p* values. The critical effect size from a sensitivity power analysis with alpha = .001, power = .80 was $\eta^2 < .0025$ ($f \leq .054$; Faul et al. 2009).

Results

Mean Item Ratings

All bi-society *r*s for the 118 mean problem item ratings were significant, with the range from .51 (Albania with Iceland) to .91 (Brazil with Argentina and Serbia). Albania had the smallest mean *r* (.62), whereas Argentina and Brazil had the largest (.81). The omnicultural mean *r* across all 14 societies was .76. These large correlations indicate that the societies were very similar with respect to which items tended to receive low, medium, or high ratings.

We next identified the 20 problem items with the highest omnicultural mean 0–1-2-ratings averaged across all 14 societies (Table 2). These 20 items include a diverse set of problems (e.g., worrying, being nervous or tense, lacking self-confidence, feeling overwhelmed by responsibilities, being forgetful, having trouble planning for the future, arguing a lot, and being stubborn, sullen, or irritable). When the 20 top ABCL items were compared with the top ASR items across 18 societies (Rescorla et al. 2016), 18 of the items were the same.

The problem items with the lowest mean ratings also appear in Table 2. These 20 items with the lowest mean ratings included such problems as self-injury, seeing or hearing things that are not there, using drugs, stealing, repeating acts over and over, and attacking people. Seventeen of the 20 items with the lowest mean ratings according to collateral reports were also among the lowest rated items according to self-reports.

Internal Consistency of Scales in Different Societies

Prior to conducting ANOVAs on scale scores, we calculated within-society alphas for all 22 scales. For each of the 14 societies, alphas for Total Problems were ≥.92, while the alphas for Internalizing and Externalizing were ≥ .83 and ≥ .85, respectively (see Table 3). Mean alphas ranged from .59 to .87 for the syndromes, *DSM*-oriented scales, Obsessive-Compulsive Problems, and Sluggish Cognitive Tempo. Mean alphas for the Personal Strengths, Friends, and Spouse/Partner scales were .76, .71, and .75, respectively. When the 22 scale alphas for each society were correlated with the 22 alphas for the other 13 societies, the mean of the bi-society *r*s was .86 (range from .79 to .90). When the 22 scale alphas for each society were correlated with the 22 alphas for the U.S. (where the scales were constructed), *r*s ranged from .72 (with Taiwan) to .94 (with Serbia), with a mean *r* of .87. These findings indicate that, despite the vicissitudes of translation, the internal consistencies of ABCL scales showed a very similar pattern across the various societies.

Problem Scale Scores

Because the 14 societies did not have equal sample sizes for all four age/gender groups, we calculated the societal means and standard deviations (*SD*s) for each scale by averaging values for the four groups. We then averaged the 14 societal means and *SD*s to obtain the omnicultural mean and mean *SD* for each scale. Table 4 displays the lowest and the highest societal mean, the omnicultural mean (and its *SD*), and the mean *SD* for each ABCL scale. For example, Total Problems scores could range from 0 to 236 but the 14 societal means ranged only from 21.1 (Taiwan) to 48.4 (Albania), with an omnicultural mean of 34.4. Ten societies had mean Total Problems scores within one *SD* (8.2) of the omnicultural mean, whereas two societies had mean Total Problems scores >1 *SD* below that mean (Taiwan and Japan) and two societies had mean Total Problems scores >1 *SD* above that mean (Lithuania and Albania). As seen in Table 4, the mean *SD* for each problem scale (e.g., 24.1 for Total Problems) was consistently larger than the *SD* of the ominicultural mean for that scale (e.g., 8.2 for Total Problems), indicating much more variance within societies than between societies in problem scores.

To explore whether there was a tendency for societies to have high Internalizing but low Externalizing scores (or vice versa), we computed a correlation between the 14 societal means for the two broad-band scales of Internalizing and Externalizing (which have no overlapping items). This *r* of .98 indicated that societies tended to have high versus low scores in general, rather than to have high Internalizing but low Externalizing scores, or the converse.

As shown in Table 5, six of the 19 problem scale societal ESs (η^2) were small (<6 %), 12 were medium (6 % -13.9 %), and one was large (≥14 %). Gender effects were significant for 15 problem scales, with the largest ES being 2 % for Rule-Breaking Behavior (M > F) and *DSM*-Anxiety Problems (F > M). Women scored higher than men on eight scales (such as Anxious/Depressed and Somatic Complaints), whereas men scored higher than women on seven scales (such as Attention Problems and Intrusive). Younger adults (ages 18–35) scored significantly higher than older adults (ages 36–59) on eight of the nine scales with significant age effects. Of the 76 interactions, 63 were not significant, while the remaining 13 were significant at $p < .001$ but had ESs ≤1 % (not displayed in Table 5), indicating strong similarity in age and gender patterns across societies.

Table 2 ABCL items with the highest and lowest omnicultural mean ratings across 14 societies

Items with Highest Omnicultural Mean Ratings[a]	Mean	Items with Lowest Omnicultural Mean Ratings	Mean
72. Worries about his/her family[b]	1.10	58. Picks skin or other parts of his/her body	.11
22. Worries about his/her future[b]	1.03	16. Cruelty, bullying, or meanness to others	.11
112. Worries a lot[b]	.78	46. Nervous movements or twitching	.11
32. Feels he/she has to be perfect [b]	.67	101. Stays away from job even when not sick or on vacation[c]	.10
3. Argues a lot [b]	.63	37. Gets in many fights[c]	.10
1. Is too forgetful [b]	.59	56d. Problems with eyes[c]	.09
45. Nervous, highstrung, or tense [b]	.56	66. Repeats certain acts over and over[c]	.08
69. Secretive, keeps things to self [b]	.56	20. Damages or destroys his/her own things[c]	.08
44. Feels overwhelmed by responsibilities [b]	.55	56 g. Vomiting, throwing up[c]	.07
99. Dislikes staying in one place for very long [b]	.55	21. Damages or destroys things belonging to others[c]	.06
19. Demands a lot of attention	.51	85. Strange ideas[c]	.06
86. Stubborn, sullen, or irritable [b]	.49	79. Speech problem[c]	.06
47. Lacks self-confidence [b]	.49	92. Does things that may cause trouble with the law[c]	.05
53. Has trouble planning for the future [b]	.47	97. Threatens to hurt people[c]	.05
78. Has trouble making decisions [b]	.47	6. Use drugs for nonmedical purposes[c]	.05
24. Does not eat well [b]	.46	84. Strange behavior[c]	.04
42. Would rather be alone than with others [b]	.45	57. Physically attacks people[c]	.04
118. Is too impatient [b]	.45	91. Talks about killing self[c]	.04
119. He/she is not good at details [b]	.43	18. Deliberately tries to hurt self or attempts suicide[c]	.03
41. Impulsive or acts without thinking	.42	70. See things that are not there[c]	.03

The omnicultural mean rating for each item was obtained by averaging the 14 mean item ratings obtained from the full sample in each society

[a] Items are listed in descending order of mean scores

[b] Parallel item was in list of 21 highest omnicultural mean ratings for the ASR

[c] Parallel item was in list of 21 lowest omnicultural mean ratings for the ASR

Personal Strengths Scale

The ABCL Personal Strengths scale includes items such as *2. Makes good use of his/her opportunities*, *4. Works up to ability*, *15. Is pretty honest*, *73. Meets responsibilities to his/her family*, *106. Tries to be fair to others*, and *123. He/she is a happy person*. Scores on the Personal Strengths scale differed considerably between societies but were quite homogeneous within societies, as indicated by the societal ES of 25 %, which was a much larger societal ES than was found for the problem scales. There was a non-significant effect for age and an ES of <1 % for gender (F > M). One society had a mean Personal Strengths score > 1 *SD* (2.0) below the omnicultural mean of 15.3 (Japan =9.8), while one society had a mean Personal Strengths score > 1 *SD* above the omnicultural mean (Flanders =17.9). For the full sample, Personal Strengths scores had a small but significant negative correlation with both Internalizing ($r = -.14$) and Externalizing ($r = -.17$).

Adaptive Functioning Scales

Friends Scale Higher scores indicate better functioning on all Adaptive Functioning scales. Scores on the Friends scale were obtained for 7,737 of the participants. The societal ES was 9 %, with Japan and Korea having the lowest mean scores and Portugal and Iceland having the highest mean scores. Women obtained slightly higher scores than men (ES < 1 %), and younger adults obtained higher scores than older adults (ES = 2 %). Age and gender effects were quite consistent across societies (interaction ESs < 1 %). The Friends scale had an r of .31 with the Personal Strengths scale. The Friends scale had a stronger negative association with Internalizing ($r = -.22$, $p < .001$) than with Externalizing ($r = -.12$, $p < .001$), a significant difference by the Hotelling-Williams z-test.

Spouse/Partner Scale The Spouse/Partner scale was completed for 4,953 participants, indicating that about 60 % of the sample had lived with a spouse/partner in the past 6 months. The societal ES was 5 %, with Hong Kong having the lowest mean and the U.S. and Flanders having the highest means. The age ES was <1 %, with younger adults having slightly higher scores than older adults. The Spouse/Partner scale was correlated .36 with Personal Strengths, .21 with the Friends scale, −.41 with Internalizing, and −.45 with Externalizing, all $p < .001$.

Table 3 Internal consistency alpha coefficients for ABCL scales across 14 societies

ABCL Scale	*N* of Items	Minimum alpha	Maximum alpha	Mean alpha
Broad-Band Scales				
Total Problems	118	.92	.97	.95
Internalizing	32	.83	.93	.89
Externalizing	35	.85	.94	.90
Syndromes				
Anxious/Depressed	14	.80	.90	.85
Withdrawn	9	.68	.85	.77
Somatic Complaints	9	.63	.81	.73
Thought Problems	9	.34	.71	.59
Attention Problems	17	.82	.90	.86
Aggressive Behavior	16	.79	.92	.87
Rule-Breaking Behavior	13	.58	.85	.76
Intrusive Behavior	6	.60	.79	.70
DSM-Oriented Scales				
Depressive Problems	15	.75	.86	.81
Anxiety Problems	6	.52	.74	.64
Somatic Problems	7	.58	.78	.67
Avoidant Personality Problems	7	.65	.81	.74
Attention Deficit Hyperactivity Problems	13	.75	.86	.82
Antisocial Personality Problems	20	.66	.88	.81
Other Scales				
Obsessive-Compulsive Problems	8	.46	.72	.59
Sluggish Cognitive Tempo	5	.53	.71	.63
Personal Strengths	11	.68	.84	.76
Friends	4	.61	.79	.71
Spouse/Partner	8	.67	.82	.75

ABCL-ASR Cross-Informant Agreement

The cross-informant agreement sample of 8,302 participants from 14 societies excluded Iceland but added China. It also excluded some participants for whom yoked ABCL-ASR data were not available, resulting in a slightly smaller *N* than for the ABCL sample.

Correlations for Scale Scores For each of the 14 societies with yoked data for the ASR and the ABCL, *r*s were calculated between scale scores on the two forms. These societal *r*s were then averaged to yield omnicultural mean cross-informant *r*s. Table 6 displays these omnicultural *r*s, their *SD*s, and the range of the 14 *r*s making up the mean. The three broad-band scales had very similar *r*s: .50 for Total Problems, .53 for Internalizing, and .50 for Externalizing, with Albania having the highest cross-informant *r*. Hong Kong had the lowest *r* for Total Problems and Internalizing, whereas China had the lowest *r* for Externalizing.

Because Internalizing and Externalizing are broad-band scales with no overlapping items that reflect two major forms of psychopathology, some additional analyses were conducted to examine them in this cross-informant sample. Although, as noted above, the omnicultural cross-informant *r*s for Internalizing and Externalizing were very similar, namely .53 (range .19–.84) and .50 (range .29–.76), we used the Asymptotic Variance *z*-test to determine whether this pattern was true in each of the 14 societies. Of the 14 *z* values, only three reached a $z \geq 1.96$, $p < .05$ and all three of these had larger Internalizing than Externalizing *r*s: Brazil: INT = .56 > EXT = .49, z = 1.96; Japan: INT = .60 > EXT = .55, z = 2.36, and Albania: INT = .84 > EXT = .76, z = 5.05).

Using multitrait-multimethod concepts (Campbell and Fiske 1959), we further examined Internalizing and Externalizing with respect to convergent versus divergent cross-informant agreement within each society. The omnicultural means for divergent *r*s (heterotrait-heteromethod) were .27 (range of .11 to .55) for ASR

Table 4 Range of society means, omnicultural means and their SDs, and mean of 14 society SDs (N = 8322)

ASR scale	Minimum mean	Maximum mean	Omnicultural mean	Omnicultural mean's *SD*	Mean *SD*
Broad-Band Scales					
Total Problems	21.1	48.4	34.4	8.2	24.1
Internalizing	6.2	13.9	9.4	2.2	7.9
Externalizing	5.1	14.5	9.3	2.7	8.4
Syndromes					
Anxious/Depressed	3.0	6.8	4.9	1.2	4.4
Withdrawn	1.8	4.8	2.6	.8	2.8
Somatic Complaints	1.2	2.4	1.9	.3	2.3
Thought Problems	.3	1.2	.8	.2	1.4
Attention Problems	4.2	8.2	5.9	1.1	5.3
Aggressive Behavior	3.0	7.4	4.8	1.3	4.8
Rule-Breaking Behavior	1.2	4.5	2.4	.9	2.8
Intrusive Behavior	1.0	3.4	2.1	.6	2.1
DSM-Oriented Scales					
Depressive Problems	2.9	5.9	4.1	.8	4.0
Anxiety Problems	2.3	6.6	4.1	1.2	2.2
Somatic Problems	.9	1.7	1.3	.2	1.8
Avoidant Personality Problems	1.6	3.3	2.3	.5	2.4
Attention Deficit Hyperactivity Problems	3.0	6.3	4.7	1.1	4.1
Antisocial Personality Problems	2.4	6.4	3.8	1.2	4.0
Other Scales					
Obsessive-Compulsive Problems	1.2	4.0	2.4	.7	2.0
Sluggish Cognitive Tempo	.8	2.6	1.6	.4	1.7
Personal Strengths	9.8	17.9	15.3	2.0	3.9
Friends	6.5	8.9	8.0	.8	2.5
Spouse/Partner	2.8	5.1	4.0	.8	2.9

Omnicultural mean = mean of the 14 society means; Mean *SD* = mean of the 14 society *SD*s

Internalizing x ABCL Externalizing and .30 (range of .12–.49) for ASR Externalizing x ABCL Internalizing, smaller than the convergent *r*s (monotrait-heteromethod) of .53 (Internalizing) and .50 (Externalizing). The largest correlations were the monomethod-heterotrait *r*s, with omnicultural means of .59 (range .44–.74) for the ASR and .59 (range .47–.71) for the ABCL. These large monomethod-heterotrait *r*s reflect the fact that Internalizing and Externalizing scores are usually positively correlated, most likely because they both reflect a person's general level of psychopathology (Caspi et al. 2014).

The 14 societies differed somewhat in the discrepancy of their convergent versus divergent *r*s. The four societies with the largest convergent *r*s had discrepancies of about .30 (e.g., Argentina, with .69 and .70 vs. .31 and .35). The five societies with moderate *r*s had somewhat smaller discrepancies (e.g., Portugal, with .40 and .44 vs. .18 and .17). The remaining five societies had the smallest discrepancies, either because all four *r*s were small (e.g., China, with .29 and .35 vs. .18 and .22) or because the *r*s were of moderate size but not very discrepant (e.g., Japan, with .55 and .60 vs. .38. and .49).

Societies also varied with respect to discrepancies between within-informant *r*s (monomethod-heterotrait) and cross-informant *r*s (monotrait-heteromethod) for Internalizing and Externalizing. For example, the within-informant *r*s between Internalizing and Externalizing for Hong Kong were .68 for the ABCL and .63 for the ASR, whereas the cross-informant *r*s for these two scales were .19 (Internalizing) and .30 (Externalizing). This pattern of much larger within-informant *r*s than cross-informant *r*s was also evident in China, Japan, Korea, Taiwan, and the USA. In contrast, for Argentina, Lithuania, Poland, and Albania, the within-informant *r*s were smaller than the cross-informant *r*s (e.g., .44 and .51 vs. .68 and .66 in Poland). For Brazil, Flanders, Portugal, and Serbia, the within-informant *r*s and cross-informant *r*s were very similar (e.g., .56 and .54 vs. .56 and .49 in Brazil).

Table 5 Significant effect sizes (η^2) for society, gender, and age on ABCL scales in 14 societies (N = 8322) and effect sizes (η^2) for informant, society, gender, and age on ABCL-ASR scales in 14 societies (N = 8,203)

	ABCL Results			ABCL-ASR Cross-Informant Results				
Scale	Society	Gender	Age	Informant	Society	Gender	Age	I x S
Broad-Band Scales								
Total Problems	9 %	ns	<1 % Y	7 %	10 %	<1	1 %	3 %
Internalizing	7 %	<1 % F	ns	6 %	8 %	2 %	ns	4 %
Externalizing	8 %	<1 % M	<1 % Y	2 %	8 %	<1 %	1 %	4 %
Syndromes								
Anxious/Depressed	7 %	1 % F	ns	6 %	9 %	3 %	<1 %	4 %
Withdrawn/Depressed	8 %	<1 % M	ns	2 %	9 %	ns	ns	3 %
Somatic Complaints	1 %	1 % F	<1 % O	3 %	2 %	2 %	ns	<1 %
Thought Problems	3 %	ns	ns	7 %	6 %	<1 %	1 %	2 %
Attention Problems	5 %	<1 % M	<1 % Y	6 %	6 %	ns	1 %	2 %
Aggressive Behavior	6 %	ns	ns	3 %	7 %	ns	<1 %	2 %
Rule-Breaking Behavior	8 %	2 % M	<1 % Y	<1 %	7 %	2 %	1 %	5 %
Intrusive Behavior	8 %	<1 % M	<1 % Y	2 %	8 %	<1 %	1 %	4 %
DSM-Oriented Scales								
Depressive Problems	4 %	<1 % F	ns	3 %	4 %	1 %	<1 %	2 %
Anxiety Problems	21 %	2 % F	ns	5 %	26 %	4 %	<1	3 %
Somatic Problems	1 %	<1 % F	ns	1 %	2 %	2 %	ns	<1 %
Avoidant Personality Problems	4 %	ns	ns	5 %	4 %	<1 %	<1 %	4 %
Attention Deficit Hyperactivity Problems	6 %	<1 % M	<1 % Y	4 %	7 %	ns	<1 %	2 %
Antisocial Personality Problems	7 %	1 % M	<1 % Y	<1 %	7 %	1 %	<1 %	4 %
Other Scales								
Obsessive-Compulsive Problems	13 %	<1 % F	ns	6 %	14 %	1 %	<1 %	2 %
Sluggish Cognitive Tempo	6 %	<1 % F	<1 % Y	11 %	6 %	1 %	< 1 %	2 %
Personal Strengths	25 %	<1 % F	ns	<1 %	36 %	ns	ns	2 %
Friends	9 %	<1 % F	2 % Y	<1 %	11 %	<1 %	3 %	<1 %
Spouse/Partner	5 %	ns	<1 % Y	<1 %	7 %	<1 %	<1 %	ns

Note.ns indicates the effect was not significant; interactions that are not shown in table were not significant or had an ES <1 %; F = females > males; M = males > females; Y = younger adults > older adults; O = older adults > younger adults. I x S = informant x society interaction

As shown in Table 6, omnicultural mean cross-informant *r*s for the 16 narrow-band problem scales (syndromes, *DSM*-scales, plus Obsessive-Compulsive Problems and Sluggish Cognitive Tempo), ranged from .37 to .52, with 14 of the 16 mean *r*s ≥ .40. Mean *r*s were .41 for the Personal Strengths scale, .58 for the Friends scale, and .60 for the Spouse/Partner scale.

Averaged across all 19 problem scales, the mean cross-informant *r* was .47, with a range from .22 for Hong Kong to .75 for Albania. Five societies had mean *r*s < .40 (Hong Kong, Korea, Taiwan, China, and the U.S.), four had mean *r*s between .40 and .49 (Flanders, Portugal, Brazil, and Serbia), three had mean *r*s between .50 and .59 (Japan, Lithuania, and Poland), and two had mean *r*s ≥ .60 (Argentina and Albania). Although within-society cross-informant *r*s varied somewhat across the 19 problem scale *r*s, the mean *r* for each society was quite an accurate reflection of the *r*s for the various scales. For example, Albania's 19 problem scales *r*s ranged from .62 to .84 (with a mean *r* of .75), indicating strong cross-informant agreement across many different kinds of problems. Agreement in Albania was also strong for Personal Strengths (.66), Friends (.76), and Spouse/Partner (.78).

In a supplementary analysis, we calculated intraclass correlations (ICCs) as an alternative measure of cross-informant agreement for a representative subset of our ABCL-ASR scales (Internalizing, Externalizing, Total Problems, Personal Strengths, Friends, and Spouse/Partner). As this was an exploratory analysis to determine whether using ICC rather than *r* yielded different results, we calculated ICCs and *r*s for our full-cross-informant sample of 8,203, rather than separately by society. For all six scales, the Pearson *r* and ICC values were identical. We therefore concluded that it would be appropriate to report only *r*s for our main analyses, consistent with much previous research.

Table 6 Cross-informant correlations (rs) for ABCL-ASR scales in 14 societies (N = 8,203)

Scale	Full SampleMean (SD)	Range
Broad-Band Scales		
Total Problems	.50 (.16)	.20–.83
Internalizing	.53 (.17)	.19–.84
Externalizing	.50 (.15)	.29–.76
Syndromes		
Anxious/Depressed	.52 (.15)	.19–.81
Withdrawn/Depressed	.46 (.15)	.25–.75
Somatic Complaints	.48 (.15)	.18–.81
Thought Problems	.37 (.13)	.16–65
Attention Problems	.49 (.14)	.20–.78
Aggressive Behavior	.47 (.16)	.24–.76
Rule-Breaking Behavior	.49 (.14)	.30–.74
Intrusive Behavior	.53 (.17)	.19–.84
DSM-Oriented Scales		
Depressive Problems	.49 (.14)	.21–.82
Anxiety Problems	.46 (.13)	.15–.68
Somatic Problems	.45 (.15)	.17–.77
Avoidant Personality Problems	.48 (.13)	.24–.73
Attention Deficit Hyperactivity Problems	.47 (.14)	.20–.75
Antisocial Personality Problems	.47 (.17)	.23–.74
Other Scales		
Obsessive-Compulsive Problems	.37 (.13)	.12–.62
Sluggish Cognitive Tempo	.40 (.13)	.14–.69
Personal Strengths	.41 (.16)	.09–.66
Friends	.58 (.13)	.32–.78
Spouse/Partner	.60 (.13)	.39–.78

To see whether rater type (spouse/partner vs. other) affected cross-informant correlations, we computed cross-informant *r*s separately for all adults rated by a spouse/partner and all adults rated by another type of rater for all 22 scales in the 12 societies that had sufficient raters of each type (not Hong Kong or Korea). After we averaged within each society the 22 *r*s from each rater type, the omnicutural means for these *r*s were .52 for spouse/partner raters and .49 for other raters, a non-significant difference by Fisher's *z*-test. Additionally, for each society, we obtained 12 difference scores, reflecting the difference for each society between the mean cross-informant *r* across 22 scales for spouse/partner raters versus other raters. Nine of these 12 difference scores were close to zero (−.02 to +.05), two were slightly larger (Serbia = .07 and Brazil = .08), and one was notably larger (China = .17), with a mean difference score across the 12 societies of .04. These findings suggest that cross-informant *r*s for spouse/partner raters versus other raters (averaged across the 22 scales) were very similar in all societies except China. Finally, to test rater type by scale rather than by society, we calculated for each of the 22 scales the omnicultural mean *r* for spouse/partner raters and for other raters separately (*r*s averaged over 12 societies). When we used Fisher's *z* transformation to test the difference between the 22 pairs of *r*s, spouse/partner *r*s were larger than *r*s from other raters for three problem scales (Thought Problems, Obsessive-Compulsive Problems, Total Problems), but lower (i.e., less favorable) for the Spouse/Partner scale.

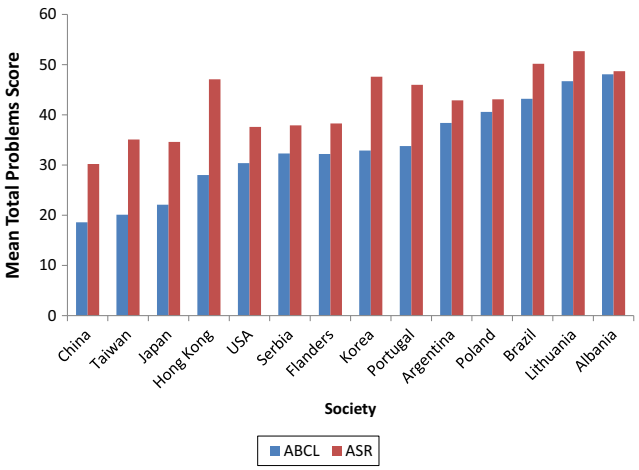

Fig. 1 Mean Total Problems scores for ABCL and ASR in 14 societies (N = 8,203)

ANOVAs for ABCL-ASR Scales As shown in Fig. 1, the mean Total Problems score was higher on the ASR than the ABCL in all 14 societies, but the size of the ASR-ABCL difference varied widely across societies. We used 2 × 14 × 2 × 2 mixed-model ANOVAs to test the effects of informant (self vs. collateral), society, gender, and age group on all ABCL and ASR problem scales, with significant ESs for main effects reported as η^2 (Table 6). Across problem scales, ASR mean problem scores were higher than ABCL mean problem scores, but informant effects varied widely across scales (from <1 % for Rule-Breaking Behavior to 11 % for Sluggish Cognitive Tempo). Informant effects also varied across societies, as shown by the significant informant x society interactions for all scales except Spouse/Partner, but ESs (<1 % to 5 %) were all small by Cohen's (1988) benchmarks. All other interactions were non-significant or ≤1 %. Four societal ESs were small (<5.9 %), 13 were medium (6 % to 13.9 %), and two were large (>13.9 %). Significant gender ESs ranged from <1 % to 2 %, except for a 4 % ES for *DSM*-Anxiety (F > M). Significant age ESs were ≤1 %.

The ANOVA on the Personal Strengths scale yielded non-significant effects of informant, gender, and age group but a very large ES for society (36 %). On the Friends and Spouse/Partner scales, informant effects were minimal (<1 %), societal ESs were medium (11 % and 7 % respectively), gender ESs were <1 %, and age ESs were 3 % and <1 %, respectively. Overall, these results show that informant (self vs. collateral), gender, and age accounted for very small portions of variance in scores on Personal Strengths and relations with friends and spouse/partner, but that scores differed considerably across societies in these domains.

To test the effect of rater type (i.e., spouse/partner vs. other) on mean ABCL and ASR scores, we redid our mixed-model ANOVAs for Total Problems, Internalizing, Externalizing, Personal Strengths, and Spouse/Partner. These informant x society x gender x age group x rater- type ANOVAs (for the 12 societies with both types of rater) yielded very similar results across the five scales, namely a rater-type effect that was significant but very small (ES ≤ 1 %). The society x rater-type interaction was significant for Total Problems and Internalizing only, with ESs of <1 %. Mean problem scores based on spouse/partner raters were slightly higher than those based on other raters for problem scales (e.g., 37.9 vs. 34.9 for Total Problems). However, for the Personal Strengths and Spouse/Partner scales, scores from spouse/partner raters indicated slightly better functioning than scores from other raters (14.4 vs. 14.0 and 4.6 vs. 3.7, respectively). Because rater-type effects on mean scores for these five scales were very small, we did not calculate them for the other 17 scales.

ABCL-ASR Mean Problem Item Ratings Agreement Within each society, we computed *Q* correlations between the 115 ABCL mean problem item ratings and the corresponding 115 ASR mean problem item ratings. The *Q* correlations were large for all 14 societies, ranging from .76 (Albania) to .98 (Argentina), with a mean = .92. This indicates that within every society there was strong agreement, on average, between adult participants and their collateral informants regarding which problem items received low, medium, or high ratings.

Dyadic *Q* Correlations for Item Ratings To examine within-dyad agreement on 0–1-2 ratings of items, we calculated separate *Q* correlations for each dyad between the 115 problem item ratings by the ASR participant and the item ratings by his/her collateral informant. When a 14 (society) × 2 (gender) × 2 (age) ANOVA was run on dyadic *Q*s (after they had been converted to Fisher's *z*s), there was an ES of 15 % for society and non-significant effects for gender and age group. Within-society mean dyadic *Q*s ranged from .28 for Hong Kong and China to .53 for Argentina, with an omnicultural mean of .39 (*SD* = .08). In all societies, *SD*s of the Total Problems *Q* correlations were large in comparison with the magnitude of the *Q*s, ranging from .15 in Korea to .22 in China (omnicultural mean of the *SD*s = .18). Thus, within every society, dyads varied quite widely in agreement on ASR-ABCL item ratings, with some dyads agreeing very well and others agreeing very poorly. Age and gender of the participants appeared to have little effect on this variation. The five societies with the lowest mean dyadic agreement were Hong Kong (*Q* = .28), China (*Q* = .28), Korea (*Q* = .31), the U.S. (*Q* = .31), and Taiwan (*Q* = .32). Seven societies had dyadic *Q*s between .38 and .48 (Flanders and Serbia = .38, Japan = .40, Lithuania = .44, Poland = .45, Brazil = .46, Portugal = .47), and two societies had dyadic *Q*s > .50 (Albania = .51, Argentina = .53).

To see whether dyadic agreement was better for spouse/partner raters than for other raters, we redid the dyadic *Q* correlations for the 12 societies separately by rater type (*N* = 7,560). An ANOVA with society, gender, age group, and rater type as factors yielded ESs of 13 % for society and non-significant effects for gender, age group, and rater type. The society x rater-type interaction was significant at $p < .001$ with an ES of <1 %. Across all 7,560 dyads, the mean dyadic *Q*s were .41 for spouse/partner raters and .40 for other raters. Mean differences for each of the 12 societies ranged from 0 (Poland = .45/.45 and Albania = .51/.51) to .04 (China = .31/.27, U.S. = .33/.29, Flanders = .40/.36). Only one society had a larger *Q* for other raters than spouse/partner raters (Japan = .39/.42). In short, dyadic *Q*s were very similar for both types of raters but ranged somewhat across societies.

Cross-Informant Agreement on Deviance Status for Total Problems Lastly, we tested ASR/ABCL agreement regarding whether the adult's Total Problems scores fell in the deviant range, defined as a score ≥ 1 *SD* above the within-society

mean (i.e., ≥ 84th percentile). Within-society cutpoints were calculated separately by gender and/or age group when societies had significant age or gender effects. Agreement, defined as both ABCL and ASR scores ≥1 *SD* above the mean (i.e., deviant) or both ABCL and ASR scores below this cutpoint (i.e., non-deviant), ranged from 74 % for Hong Kong to 88 % for Albania (omnicultural mean = 81 %). Dyads thus disagreed on deviance status for 12 % to 26 % of cases across the societies. The kappa statistic was also used to measure cross-informant agreement, with Hong Kong showing the lowest kappa (.08) and Albania the highest (.58). The omnicultural mean kappa of .32 indicated modest cross-informant agreement across the 14 societies. It should be noted that Hong Kong also had the lowest scores and Albania also had the highest scores on the other four decision statistics reported below; for purposes of labeling these statistics with their traditional names, the ASR Total Problems score was used as the criterion for deviance.

In all societies, most adults whose ASR Total Problems scores were in the non-deviant range also had ABCL Total Problems scores in the non-deviant range (from 85 % to 91 %, mean specificity = 89 %). Additionally, most adults whose ABCL Total Problems scores were in the non-deviant range had ASR Total Problems scores in the non-deviant range (from 84 % to 93 %, mean negative predictive value = 89 %). When the ASR Total Problems score was in the deviant range, the ABCL Total Problems score was in the deviant range on average for less than half the dyads (from 24 % to 66 %, sensitivity mean = 43 %). Similarly, when the ABCL Total Problems score was in the deviant range, the ASR Total Problems score was in the deviant range on average for less than half the dyads (from 24 % to 66 %, positive predictive value mean = 42 %). It was thus common in all societies both for collaterals to not corroborate deviance based on adults' self-reports and for adults to not corroborate deviance based on collaterals' reports. However, non-corroboration rates varied widely across the 14 societies.

Discussion

ABCL Findings in 14 Societies

Because very few studies of collateral reports about adults' mental health problems have been published, our study of collateral reports by 8322 adults in 14 very different societies represents an important advance. It should be noted that, even though the target adults in many of our samples were recruited via random sampling, collateral informants nominated by those target adults constitute at best a quasi-random sample or what might be termed a "second-order" random sample. Unlike a more typical epidemiological sample, the sampling frame for our study comprised collaterals nominated by samples of adults 18 to 59 years of age living in 14 societies. Our findings therefore generalize only to those populations. Moreover, the conclusions drawn from our results apply only to the societies we studied. That is, we do not attempt to generalize our findings to societies not included in our analyses.

The same ABCL problem items tended to obtain relatively low, medium, or high ratings in all societies, as manifested by an omnicultural mean *r* of .76 across the 14 societies, virtually identical to the *r* of .75 that Rescorla et al. (2016) reported across 17 societies for the ASR. Furthermore, as Table 2 shows, the most commonly endorsed items were the same on the ASR and ABCL (e.g., worrying, being nervous or tense, lacking self-confidence, feeling overwhelmed by responsibilities). When considered individually, these problems are not necessarily symptoms of serious psychopathology. However, people seeking professional help often do so because they experience multiple problems of this kind. The least commonly endorsed problem items were also very similar for the ABCL and the ASR. The ominicultural mean *r* of .86 between scale alphas further demonstrates that the items performed similarly in different societies.

With respect to scale scores, similarities across societies were also found with respect to ABCL age and gender patterns, as seen in minimal societal interactions with age and gender. Consistent with our ASR findings (Rescorla et al. 2016), women scored higher than men on ABCL internalizing-type scales (such as Anxious/Depressed and Somatic Complaints), whereas men scored higher than women on ABCL externalizing-type scales (such as Rule-Breaking Behavior and Intrusive). Younger adults (ages 18–35) scored significantly higher than older adults (ages 36–59) on eight of the nine scales that had significant age effects, also consistent with ASR findings. Identity of the collateral informant (spouse/partner vs. other) had minimal effects on ABCL scores, based on ANOVAs for Total Problems, Internalizing, Externalizing, Personal Strengths, and Spouse/Partner.

Our ABCL analyses also revealed some important differences between societies. For example, ABCL mean Total Problems scores for the 14 societies ranged from 21.1 for Taiwan to 48.4 for Albania. However, 10 societies had quite similar scores (i.e., within one *SD* of the omnicultural mean), including Argentina, Flanders, Poland, Albania, Korea, and the U.S. These societies differ not only in race/ethnicity, geography, political system, economic status, and religion but also along Hofstede's (2011) cultural dimensions (individualism/collectivi, uncertainty avoidance, power distance, masculinity/femininity, long/short-term orientation, and indulgence/restraint). The societal differences found in ABCL scale scores do not seem well explained by conventional categories such as

individualist/collectivist, East/West, Confucian/European, or rich/poor. For example, two Asian societies (Taiwan and Japan) scored >1 *SD* below the ABCL omnicultural mean for Total Problems, but two other Asian societies (Hong Kong and Korea) did not. Additionally, two former East Bloc societies (Lithuania and Albania) scored >1 *SD* above the omnicultural mean for Total Problems, but two others (Serbia and Poland) did not.

Striking societal differences were found on the Personal Strengths scale. Whereas most societal effects on problem scales were small to medium, the societal effect on Personal Strengths was very large (25 %) and the within-society variation (*SD* = 3.9) was smaller than for the problem scales (a pattern also found on the ASR). Japan and Korea had low mean Personal Strengths scores on both the ASR and the ABCL, suggesting a need for emic investigation (Pike 1967).

ABCL-ASR Cross-Informant Agreement Findings

Cross-informant agreement between self-reports and collateral reports for adult psychopathology has received relatively little research attention. Furthermore, to our knowledge, no international comparisons of cross-informant agreement regarding adult psychopathology have been published. Our study, therefore, addresses a significant gap in the literature.

Our cross-informant analyses revealed many similarities across societies. ASR mean scores were significantly higher than ABCL mean scores for most problem scales in most societies. The informant effect varied somewhat across societies but the ESs for informant x society interactions were small (<1 % to 5 %). For the five scales analyzed, rater-type had very small effects (≤1 %) on scores, indicating that spouse/partner ratings yielded very similar scores to those obtained from other kinds of raters. Additionally, *Q* correlations between ASR and ABCL mean problem item ratings were large in every society (range = .76 to .98, mean = .92), indicating great consistency across societies in how well, on average, collaterals and assessed adults agreed on which problem items they rated low, medium, or high. Furthermore, with great consistency across societies, collaterals and the assessed adults agreed well on deviance status (from 74 % to 88 % of cases across societies, mean = 81 %), consistent with findings for parents and their adolescent children (Rescorla et al. 2013). In each society, disagreements were about evenly divided between collateral non-corroboration of self-reported deviance and assessed adult non-corroboration of collateral-reported deviance.

An important implication of these dichotomous results regarding non-corroboration is that using an "AND" rule to establish deviance (i.e., requiring both informants to agree the target individual is deviant) yields much lower "prevalence" than using an "OR" rule (i.e., either informant says the target individual is deviant), an issue discussed by Youngstrom Calabrese, and Findling (2003). With our non-corroboration rates averaging 42–43 % in both directions, it is clear that many more people would qualify as deviant if deviance status is based on one informant only, rather than requiring both informants to report deviance.

We found only minor differences between ABCL-ASR *r*s obtained for spouse/partner versus other raters, with the mean difference very close to zero in nine of the 12 societies analyzed, but higher in China, which had *r*s of .46 versus .30. For *r*s transformed to Fisher's *z*s, differences between spouse/partner versus other raters were significant for only three of 22 scales (Thought Problems, Total Problems, and Spouse/Partner). Thus, contrary to what might be expected, cross-informant agreement was not markedly better for spouse/partner raters than for other types of raters.

Results for 14 societies (and rater types pooled) yielded a mean cross-informant *r* averaged across all 19 problem scales of .47, quite close to the mean *r*s Achenbach et al. (2005) found between adults' self-ratings and collaterals' ratings and (.43 for internalizing problems, .44 for externalizing problems, .45 for all problems) in their meta-analysis of 108 studies. Our mean *r* also approximated the mean *r* of .41 that Rescorla et al. (2013) found between adolescent self-ratings and parent ratings in 25 societies. However, our cross-informant analyses revealed some important societal differences, as did the Rescorla et al. (2013) study of parent-adolescent agreement. Five societies had mean *r*s < .40 (Hong Kong, Korea, Taiwan, China, and the U.S.), two had mean *r*s ≥ .60 (Argentina and Albania), and the other seven had mean *r*s between .40 and .59. Although four of our five Asian samples had among the lowest cross-informant *r*s for scale scores, the U.S. was also in this low *r* group but Japan was not. The 14 societies we analyzed also differed widely in their mean dyadic *Q* correlation for item ratings (from .28 to .53, mean = .39). However, the societies were similar in all having large within-society *SD*s for dyadic *Q*, indicating that dyads within every society varied widely in item agreement. These findings are similar to those reported by Rescorla et al. (2013) for agreement between item ratings by adolescents and their parents.

Limitations and Conclusions

Limitations of our study include that we could not test associations of ABCL scores with urban/rural residence, religion, or ethnicity; that convenience sampling and low response rates might have affected some ABCL samples; and that not all ABCL participants had a yoked ASR target participant. Additionally, the omnicultural mean alphas for two of the problem scales were < .60 and the minimum alphas for some scales were rather low. However, the mean of the bi-society *r*s for alphas was .86 (range from .79 to .90), indicating considerable

consistency across the 14 societies with respect to the pattern of internal consistencies across the 22 ABCL scales.

Because ours was an etic study, the same instrument was used in every society. Research employing an emic approach, which would utilize items designed to address culture-specific constructs, might yield different results. For example, Cheung et al. (2003) summarize research pursuing the goal of an "indigenization movement in Asian psychology." This movement has led Asian researchers in many countries to construct personality instruments that they hypothesize assess characteristics unique to their specific cultures. As Cheung et al. (2003) note, such instruments frequently include a dimension regarding the "social nature of the self and the person in relational contexts," which Asian psychologists consider to be lacking in many Western personality inventories. Because the ABCL and ASR include Friends and Spouse/Partner scales, we were able to compare societal differences on these social aspects of adaptive functioning. Interestingly, Japan and Korea had the lowest scores on the Friends scale and Hong Kong had the lowest scores on the Spouse/Partner scale.

Despite the limitations noted, our study has numerous strengths. Sample sizes were large and many samples were representative of their populations; the 14 societies differed widely in economic, political, religious, and ethnic characteristics; use of the same instrument in all societies allowed direct statistical tests of societal differences; and we had yoked cross-informant samples totaling >8,000 adults and collaterals from 14 diverse societies. As with the ASR (Rescorla et al. 2016), strong consistency across societies was found in correlations between mean item ratings and in age and gender differences, but significant differences between societies were found in scale scores, with a particularly large ES for the Personal Strengths scale. We also found interesting and important associations between problem scores, and adaptive functioning.

Our cross-informant agreement findings were very similar to those reported by Rescorla et al. (2013) for parents and their adolescent children, namely strong similarity across societies or informant effects on problem scores (problem scores from self-ratings higher than those from collaterals' ratings), as well as for cross-informant agreement on mean item ratings, within-society variation in dyadic agreement for item ratings, and decision statistics on deviance status. Societal differences found for the ASR/ABCL were also very similar to those reported by Rescorla et al. (2013) for the YSR/CBCL, namely that societies varied substantially in cross-informant *r*s for problem scales and in mean dyadic *Q* correlations for item ratings.

Research and Clinical Applications

Our study indicates that epidemiological data on mental health problems can be obtained from collateral informants at relatively low cost when indigenous investigators use a standardized assessment instrument that does not require professional time for administration or scoring. Using this methodology, we were able to obtain data from numerous societies with turbulent recent histories and limited economic resources, such as Albania, Latvia, and Serbia. Furthermore, collateral report data were easily obtained in Asian societies, including Japan, Taiwan, Hong Kong, and Korea, as well as in European and South American societies. Because the same assessment instrument was used in all societies (following rigorous translation and back-translation), and the data comprised quantitative ratings, they could be easily merged to enable international comparisons. Large correlations between mean item ratings across societies indicated that ABCL items were interpreted quite similarly in the 14 different societies. The list of items with the highest and lowest mean ratings should be useful for researchers who assess mental health problems in different societies.

Our findings also have important implications for clinical practice. Because all ABCL scales have counterparts on the ASR, their combined use efficiently assesses many different aspects of a person's mental health as seen from different perspectives. Interestingly there were only minor differences in cross-informant agreement when spouse/partner raters were compared with other kinds of raters. The significant associations found between the Spouse/Partner scale and both Internalizing and Externalizing problem scores highlight the intertwined nature of mental health problems and adaptive functioning in daily life. Consistent with our ASR findings, societal ESs were rather modest for most problem and adaptive functioning scales, but the ES was very large for the Personal Strengths scale. This suggests that cultural factors may have stronger effects on adults' self- and collateral-reported positive qualities than on their reports of problems, an important consideration for clinicians treating clients from different cultural backgrounds.

Compliance with Ethical Standards

Conflict of Interest The ABCL and the ASR are published by the nonprofit University of Vermont Research Center for Children, Youth, and Families, from which authors Rescorla, Achenbach, Ivanova, and Turner receive remuneration.

Experiment Participants In each society, conventions for obtaining informed consent required by the investigator's research institution were followed.

References

Achenbach, T. M. & Rescorla, L. A. (2003). *Manual for the ASEBA Adult Forms & Profiles*. Burlington: University of Vermont, Research Center for Children, Youth, and Families.

Achenbach, T. M. & Rescorla, L. A. (2015). *Multicultural supplement to the ASEBA Adult Forms & Profiles*. Burlington: University of Vermont, Research Center for Children, Youth, and Families.

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., Krukowski, R. A., Dumenci, L., & Ivanova, M. Y. (2005). Assessment of adult psychopathology: Meta-analyses and implications of cross-informant correlations. *Psychological Bulletin*, *131*, 361–382. doi:10.1037/0033-2909.131.3.361.

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

Au, A., & Leung, P. (2012). [Hong Kong ABCL-ASR data]. Unpublished raw data.

Caldas, J. C. (2012). [Portuguese ABCL-ASR data]. Unpublished raw data.

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105.

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., et al. (2014). The p factor one general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, *2*(2), 119–137.

Cheung, F. M., Cheung, S. F., Wada, S., & Zhang, J. (2003). Indigenous measures of personality assessment in Asian countries: A review. *Psychological Assessment*, *15*, 280–289. doi:10.1037/1040-3590.15.3.280.

Chen, Y. C. (2012). [Taiwanese ABCL-ASR data]. Unpublished raw data.

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

Decoster, J., & Fontaine, J. R. J. (2012). [Flemish ABCL-ASR data]. Unpublished raw data.

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:10.1080/15374416.2011.533405.

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., Bunnell, B. E., & Beidel, D. C. (2013a). Informant discrepancies in adult social anxiety disorder assessments: Links with contextual variations in observed behavior. *Journal of Abnormal Psychology*, *122*, 376–386. doi:10.1037/a0031150.

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

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D., et al. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, *141*, 858–900. doi:10.1037/a0038498.

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.

Ellis, B. B. & Kimmel, D. D. (1992). Identification of unique cultural response patterns by means of item response theory. *Journal of Applied Psychology*, *77*, 177–184.

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.

Funabiki, Y. (2012). [Japanese ABCL-ASR data]. Unpublished raw data.

Guðmundsson, H. S., & Árnadóttir, H. A. (2012). [Icelandic ABCL-ASR data]. Unpublished raw data.

Hofstede, G. (2011). Dimensionalizing Cultures: The Hofstede Model in Context. *Online Readings in Psychology and Culture*, *2*(1), 1–26. doi:10.9707/2307-0919.1014.

Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Turner, L. V., Árnadóttir, H. A., Au, A., et al. (2015a). Syndromes of collateral-reported psychopathology for ages 18-59 in 18 Societies. *International Journal of Clinical and Health Psychology*, *15*, 18–28. doi:10.1016/j.ijchp.2014.07.001.

Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Turner, L. V., Au, A., Bellini, M., et al. (2015b). Syndromes of self-reported psychopathology for ages 18-59 in 29 societies. *Journal of Psychopathology and Behavioral Assessment*, *37*, 171–183. doi:10.1007/s10862-014-9448-8.

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

Liu, J. (2012). [Chinese ABCL-ASR data]. Unpublished raw data.

Markovic, J. (2012). [Serbian ABCL-ASR data]. Unpublished raw data.

Meyer, G.J., Finn, S.E., Eyde, L.D., Kay, G.G., Moreland, K.L., Dies, R.R., …Reed, G.M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. American Psychologist, 56, 128–165.

Oh, K. J. & Kim, Y. A. (2010). *Manual for the Korean Version of the ASEBA Adult Forms and Profiles*. Seoul: Huno Press.

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

Rescorla, L.A., Ginzburg, S., Achenbach, T.M., Ivanova, M.Y., Almquist, F., Begovac, I., Bilenberg, N.,…. 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

Rescorla, L.A., Achenbach, T.M., Ivanova, M.Y., Turner, L.V., Althoff, R.R., Au, A. …. Zasepa M. (2016). Problems and adaptive functioning reported by adults in 17 societies. *International Perspectives in Psychology*: *Research*, *Practice*, *Consultation*, in press.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. *In* Berkowitz, L. Advances in experimental social psychology 10. New York: Academic Press. *pp.* 173–220. ISBN 0–12-015210-X

Samaniego, V. C. & Vázquez, N. (2012). *Adult psychopathology: Is there any agreement between self-reports and reports by other informants? Presented at the 30th International Congress of Psychology.* South Africa: Cape Town.

Silvares, E., & da Rocha, M. M. (2012). [Brazilian ABCL-ASR data]. Unpublished raw data.

Šimulionienė, R., Brazdeikienė, L., Rugevičius, M., Gedutienė, R., & Žakaitienė, A. (2010). The psychometric properties of the Lithuanian version of ASEBA adult forms. *Psichologija*, *42*, 23–43.

Sokoli, E. (2013). [Albanian ABCL-ASR data]. Unpublished raw data.

Stephenson, W. (1935). Technique of factor analysis. *Nature*, *136*, 297.

Tomlinson, M., Rudan, I., Saxena, S., Swartz, L., Tsai, A. C., & Patel, V. (2009). Setting priorities for global mental health research. *Bulletin of World Health Organization*, *87*, 438–446. doi:10.2471/BLT.08.054353.

World Health Organization World Mental Health Survey Consortium (2004). Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization world mental health surveys. *Journal of the American Medical Association*, *291*, 2581–2590. doi:10.1001/jama.291.21.2581.

Youngstrom, E. A., Findling, R. L., & Calabrese, J. R. (2003). Who are the comorbid adolescents? Agreement between psychiatric diagnosis, youth, parent, and teacher report. *Journal of Abnormal Child Psychology*, *31*(3), 231–245.

Zasepa, E. (2012). [Polish ABCL-ASR data]. Unpublished raw data.