


Behavioral/Emotional Problems of Preschoolers: Caregiver/Teacher Reports From 15 Societies

Journal of Emotional and Behavioral Disorders
20(2) 68–81
© Hammill Institute on Disabilities 2012
Reprints and permission: <http://www.sagepub.com/journalsPermissions.nav>
DOI: 10.1177/1063426611434158
<http://jebd.sagepub.com>


Leslie A. Rescorla, Thomas M. Achenbach, Masha Y. Ivanova, Niels Bilenberg, Gudrun Bjarnadottir, Silvia Denner, Pedro Dias, Anca Dobrea, Manfred Döpfner, Alessandra Frigerio, Miguel Gonçalves, Halldór Guðmundsson, Roma Jusiene, Solvejg Kristensen, Felipe Lecannelier, Patrick W. L. Leung, Jianghong Liu, Sofia P. Löbel, Bárbara César Machado, Jasminka Markovic, Paola A. Mas, Elaheh Mohammad Esmaeili, Rosario Montiroso, Julia Plück, Adelina Ahmeti Pronaj, Jorge T. Rodriguez, Pamela O. Rojas, Klaus Schmeck, Mimoza Shahini, Jaime R. Silva, Jan van der Ende, and Frank C. Verhulst

Abstract

This study tested societal effects on caregiver/teacher ratings of behavioral/emotional problems for 10,521 preschoolers from 15 societies. Many societies had problem scale scores within a relatively narrow range, despite differences in language, culture, and other characteristics. The small age and gender effects were quite similar across societies. The rank orders of mean item ratings were similar across diverse societies. For 7,380 children from 13 societies, ratings were also obtained from a parent. In all 13 societies, mean Total Problems scores derived from parent ratings were significantly higher than mean Total Problems scores derived from caregiver/teacher ratings, although the size of the difference varied somewhat across societies. Mean cross-informant agreement for problem scale scores varied across societies. Societies were very similar with respect to which problem items, on average, received high versus low ratings from parents and caregivers/teachers. Within every society, cross-informant agreement for item ratings varied widely across children. In most respects, results were quite similar across 15 very diverse societies.

Keywords

international, cross-cultural, preschoolers, teacher-reported problems, cross-informant agreement

Psychopathology in preschool children has received increased research and clinical attention since about 1990 (Carter, 2010). Because preschoolers in many societies attend day care or preschool programs, caregivers/teachers throughout the world are potentially important sources of information about preschoolers' behavioral/emotional problems. One can hypothesize that differences between societies in cultural values, language, ethnicity, standard of living, religion, and other characteristics might lead to major societal differences in caregiver/teacher reports about preschoolers' problems. However, to test this hypothesis, international *etic* studies (Pike, 1967) are needed to compare standardized assessment data from many different societies. As defined by Pike, *etic* studies apply the same standardized procedures for assessing people in different societies.

Relatively little is known about the societal effects on caregiver/teacher reports of preschoolers' problems. Few studies have tested societal differences or societal interactions with gender and age on preschoolers' problems reported by caregivers/teachers. Nor is much known about similarities or differences between societies with respect to which behavioral/emotional problems are reported most often by caregivers/teachers. With respect to cross-informant agreement, little is known about societal differences in problems reported by parents versus caregivers/teachers.

Corresponding Author:

Leslie A. Rescorla, Bryn Mawr College, 101 N. Merion Avenue, Bryn Mawr, PA, 19010, USA
Email: lrescorl@brynmawr.edu

Effects of Society on Caregiver/Teacher Reports of Preschoolers' Problems

In one of the few published studies of preschool teachers' reports from multiple societies, LaFreniere et al. (2002) reported within-society analyses of teachers' ratings on the 30-item *Social Competence and Behavior Evaluation Inventory* (SCBE-30; LaFreniere & Dumas, 1995) for 4,640 children from eight societies (the United States, Austria, Brazil, Quebec, China, Italy, Japan, and Russia). Principal components analysis yielded the same three factors in all societies (social competence, anger aggression, and anxiety withdrawal), with some relatively minor differences between societies on item factor loadings. Cronbach's alphas were also comparable across the societies. Within-society analyses of age and gender effects yielded some similarities across societies (e.g., social competence scores higher for older children and girls than for younger children and boys; anger-aggression scores higher for boys than girls) but also some differences (e.g., age effects significant in only a few societies for anger aggression and anxiety withdrawal). Although LaFreniere et al. provided important international findings on caregiver/teacher reports of children's problems, a limitation was that data from the different societies were not combined into analyses that directly tested societal effects.

Findings for ratings of preschoolers' behavioral/emotional problems on the 99-item *Caregiver-Teacher Report Form* (C-TRF; Achenbach & Rescorla, 2000) have also been reported for several societies. Achenbach and Rescorla (2000) reported normative data for the C-TRF based on findings for 1,192 U.S. preschoolers. Subsequently, C-TRF findings were reported by Frigerio et al. (2006) for 526 Italian preschoolers; by Kristensen, Henriksen, and Bilenberg (2010) for 624 Danish preschoolers; by Liu, Cheng, and Leung (2010) for 876 Chinese preschoolers; and by Denner and Schmeck (2005) for 1,050 German preschoolers. Findings from these studies indicated some similarities but also some differences across the 5 societies in caregiver/teacher reports of preschoolers' problems. However, because the data from the 5 societies were not combined, societal effects were not directly tested. Ivanova et al. (2011) used confirmatory factor analysis (CFA) to test the fit of data from 14 societies to the C-TRF problem syndrome structure that had previously been derived from a mainly U.S. sample. Although Ivanova et al. reported generally good model fit in all 14 societies, data for each society were analyzed separately rather than being aggregated.

To our knowledge, C-TRF data from many societies have hitherto not been aggregated into a single data set so that societal effects could be directly tested. This methodology has been used, however, by Rescorla et al. (2011) for testing societal effects on parents' reports about preschoolers'

problems based on ratings obtained in 24 societies with the 99-item *Child Behavior Checklist for Ages 1½-5* (CBCL/1½-5; Achenbach & Rescorla, 2000). Rescorla et al. found small to medium societal effect sizes (ESs) on problem scores (3%–12%) and very small gender, age, and interaction effects (<1%). To measure the degree to which particular problem items tended to receive high, medium, or low ratings in the 24 different societies, Rescorla et al. computed the mean of the item ratings received for each of the 99 problem items. Large correlations of .63 to .94 were found between mean item ratings for every pair of societies, indicating considerable similarity in the problems that received high, medium, or low ratings.

Effects of Society on Cross-Informant Agreement

Modest cross-informant agreement is one of the most robust phenomena in developmental psychopathology (Achenbach, McConaughy, & Howell, 1987; De Los Reyes & Kazdin, 2005). For example, Achenbach et al. (1987) reported a mean correlation (r) of only .28 between reports by parents and teachers. However, few studies have examined cross-informant agreement for preschool children, whether in the United States or abroad.

Gross, Fogg, Garvey, and Julion (2004) examined agreement between parents and caregivers/teachers for a sample of 241 children ages 2 to 4 years from low-income families attending day care programs in the United States. The correlation between externalizing behavior problems rated by caregivers/teachers on the C-TRF and by parents on the 36-item *Eyberg Child Behavior Inventory* (ECBI; Eyberg & Pincus, 1999) was .17. Also in the United States, Winsler and Wallace (2002) reported that parents' and teachers' ratings of 47 preschool children on the 76-item *Preschool and Kindergarten Behavior Scales* (Merrell, 1994) were "somewhat more in agreement" for externalizing problems ($r = .37$) than internalizing problems ($r = .23$). Parents rated their children as having significantly more internalizing and externalizing problems than did teachers. More recently, Crane, Mincic, and Winsler (2011) reported r s of .20 to .28 between parent and caregiver/teacher ratings on the *Devereux Early Childhood Assessment* (DECA; LeBuffe & Naglieri, 1999) for 7,756 ethnically diverse U.S. preschoolers from poor families.

Cross-informant agreement between parent and caregiver/teacher ratings of preschoolers on the CBCL/1½-5 and the C-TRF has been reported for a few societies (Achenbach & Rescorla, 2000; Frigerio et al., 2006; Liu et al., 2010), with mean cross-informant r s for problem scale scores ranging from .18 in China to .40 in the United States. However, to our knowledge, no previous studies have directly tested societal effects on cross-informant agreement for preschoolers. Thus, societies need to be compared with respect to whether

Table 1. Characteristics of C-TRF Samples From 15 Societies

Society ^a	Reference	<i>n</i>	Age (in years)	Male (%)	Sample
Austria	Schmeck and Skrabel (2004)	342	3–5	52	Regional school-based
Chile	Lecannelier et al. (2011)	848	1½–5	54	Regional school-based
China	Liu et al. (2010)	931	4–5	53	Regional school-based
Denmark	Kristensen et al. (2010)	625	1½–5	49	Regional hospital birth cohort
Germany	Denner and Schmeck (2005); Döpfner and Plück (2009)	1,237	2–5	52	Regional school-based
Iceland	Guðmundsson and Bjarnadóttir (2009)	299	1½–5	49	National school-based
Iran	Mohammad Esmali (2009)	1,350	1½–5	49	National school-based
Italy	Frigerio et al. (2006)	526	1½–5	52	Regional school-based
Kosovo	Shahini and Pranvera (2010)	322	1½–5	52	Regional school-based
Lithuania	Jusiene et al. (2007)	824	1½–5	53	National school-based
The Netherlands	Tick et al. (2007)	371	1½–5	52	Regional household
Portugal	Dias et al. (2009)	384	1½–5	51	Regional school-based
Romania	Dobrea (2009)	893	2–5	47	National school-based
Serbia	Markovic (2011)	377	4–5	43	Regional school-based
The United States	Achenbach and Rescorla (2000)	1,192	1½–5	51	National household; hospital birth cohort

Note: C-TRF = *Caregiver-Teacher Report Form*.

^aData from the samples participating in the present study were also used by Ivanova et al. (2011), but analyses in the two studies did not overlap. The Romanian sample used in the present study was slightly smaller than that used by Ivanova et al. (2011), because children with anomalies in age or gender scores were excluded from the present report.

parents and caregivers/teachers typically report similar levels of problems, and whether this differs by type of problem or by the age and gender of the child. Moreover, societal differences need to be tested for cross-informant correlations and for agreement regarding which problems tend to receive high, medium, or low ratings.

Purposes of the Present Study

The present study had two main purposes. The first purpose was to test societal effects on caregiver/teacher ratings of preschoolers' problems. The second purpose was to test societal effects on cross-informant agreement between parents and caregivers/teachers. To achieve the first purpose, we tested societal effects on preschoolers' problem scores obtained from caregivers/teachers' ratings of 10,521 children ages 1½ to 5 years in 15 societies. Specifically, we used analyses of covariance (ANCOVA) to test the effects of society and gender on mean problem scores (with age as a covariate) and correlations to test societal effects on mean item ratings and scale internal consistencies. To achieve the second purpose, we tested societal effects on agreement between parents' and caregivers/teachers' ratings of 7,380 preschoolers in the 13 societies for which both kinds of ratings were available for the same children. Specifically, we used ANOVAs to test effects of society, informant, and

gender on mean problem scale scores, separately for younger and older preschoolers, plus correlations to test for societal differences in cross-informant agreement for scale scores and item ratings.

Method

Samples

Data were collected by indigenous investigators interested in studying preschoolers' behavioral/emotional problems as reported by caregivers/teachers in their own societies. These investigators shared the data they had collected with the first author for multicultural comparisons. In each society, data were obtained from general population samples broadly representative of their societies, with the sampling frame being national in 5 societies and regional in 10 societies. Recruitment was done through schools in 12 societies, through a hospital birth cohort in Denmark, through households in the Netherlands, and through households in the United States supplemented with a birth cohort of children born in 31 U.S. hospitals. In each society, sampling was based on various demographic factors, though details of the sampling varied across societies. Data from the 14 non-U.S. societies were used in Ivanova et al.'s (2011) CFA study, but the analyses do not overlap between the two studies.

Table 1 summarizes sample information from the 15 societies, with sample sizes ranging from 299 for Iceland to 1,350 for Iran ($N = 10,521$). Boys comprised from 43% to 54% of the samples. The total sample had the following age composition: ages 1½ to 2 years (4%), age 2 years (11%), age 3 years (16%), age 4 years (33%), and age 5 years (36%). Two samples (Germany and Romania) had no children younger than 2 years, two samples (China and Serbia) had no children in the 1½ to 3 age group, and Austria had only two children younger than 4 years of age.

Children who had been referred for mental health or special education services in the preceding 12 months were excluded from the Italian sample before we received the data. Information about referral was not available for most of the other 14 samples. Because referral rates for preschoolers tend to be low and preschool mental health services are scarce in most societies, it is unlikely that many of the children had been referred for services.

In each society, conventions for obtaining informed consent required by the investigator's research institution were followed. Cases were identified only by numerical codes. Children were excluded if ratings were missing for more than eight problem items, with no cases excluded for six societies, <1% of cases excluded for five societies, 2% excluded for one society, and 6% to 7% excluded for three societies.

To examine comparability across societies in cross-informant correlations between parents and caregivers/teachers, a cross-informant data set was created for each society except Chile and Kosovo, for which parallel parent and caregiver/teacher ratings were not available. For all cross-informant analyses, cases without both parent data and caregiver/teacher data were excluded, reducing the sample size to 7,380 children having both forms.

Measures

The primary measure used in this study was the C-TRF (Achenbach & Rescorla, 2000), which obtains ratings from caregivers and teachers who observe children ages 1½ to 5 years in preschools, early childhood programs, or day care settings. The C-TRF has 99 specific problem items, all of which are rated on a 3-point scale (0 = *not true [as far as you know]*; 1 = *somewhat or sometimes true*; 2 = *very true or often true*) based on children's functioning over the preceding 2 months. For cross-informant analyses, parent ratings were obtained on the CBCL/1½–5 (Achenbach & Rescorla, 2000), which also has 99 items rated on a 3-point scale (0–1–2) based on children's functioning over the preceding 2 months. There are 82 items in common across the two forms.

Six syndromes of co-occurring problems were identified for the CBCL/1½–5 and C-TRF through exploratory and CFA of item ratings (Achenbach & Rescorla, 2000). The

0–1–2 ratings of items comprising each scale are summed to yield a raw scale score.

Second-order factor analyses of the six syndromes yielded two broad-band groupings: *Internalizing* (composed of the Emotionally Reactive, Anxious/Depressed, Somatic Complaints, and Withdrawn syndromes; maximum possible score on the C-TRF = 64) and *Externalizing* (composed of the Attention Problems and Aggressive Behavior syndromes; maximum possible score on the C-TRF = 68). The Total Problems scale is the sum of the 0–1–2 ratings on all problem items, with a maximum possible score of 198. The C-TRF and the CBCL/1½–5 have five *Diagnostic and Statistical Manual of Mental Disorders (DSM)*-oriented scales (Affective Problems, Anxiety Problems, Pervasive Developmental Problems, Attention-Deficit/Hyperactivity Problems, and Oppositional Defiant Problems), comprising items identified by an international panel of experts as being very consistent with diagnostic categories of the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*; American Psychiatric Association, 1994).

Caregivers/teachers in the 14 non-U.S. societies completed C-TRFs translated into the language of their society. Foreign language versions were created by translators fluent in English and the foreign language in question. Simple language was used to be consistent with the fifth-grade reading level of the U.S. version. To verify that translations captured the original meanings, independent back translations into English were done, which then guided fine tuning of the translation in an iterative process. Translations were also used for the CBCL/1½–5, as reported by Rescorla et al. (2011).

Investigators in each society provided 0–1–2 ratings of the 99 specific problem items for each child. These ratings were used to compute raw scale scores for the C-TRF's six syndromes, five *DSM*-oriented scales, and three broadband scales (Internalizing, Externalizing, and Total Problems), as well as for a 7-item Stress Problems scale (Achenbach & Rescorla, 2010), which was derived from research with preschoolers who varied in their exposure to traumatic events. The Stress Problems scale includes items such as 5 = *can't concentrate, can't pay attention for long*; 47 = *nervous, high-strung, or tense*; and 82 = *sudden changes in mood or feelings*.

Data Analysis

As is typical for problem scores in general population samples where most children have relatively few problems, scale scores were positively skewed. However, as the general linear models we used are very robust with respect to deviations from normality, especially with large samples manifesting the same skew pattern (Kirk, 1995), we analyzed untransformed raw scores. Because the large samples provided such high statistical power that even very small

Table 2. Ranges, Omnicultural Means, Standard Deviations, and Alphas Across 15 Societies

C-TRF scale (maximum possible score)	Range ^a	Omnicultural <i>M</i> (<i>SD</i>) ^b	Omnicultural <i>SD</i>	Mean α ^c
Total Problems (198)	10.6–37.8	24.1 (8.6)	20.2	.95 (99)
Internalizing (64)	3.0–12.0	7.3 (2.8)	6.7	.88 (32)
Externalizing (68)	5.4–15.2	10.3 (3.3)	10.4	.94 (34)
Emotionally reactive (14)	0.7–2.8	1.6 (0.7)	1.9	.68 (7)
Anxious/Depressed (16)	1.1–3.7	2.4 (0.8)	2.4	.73 (8)
Somatic Complaints (14)	0.2–2.4	0.8 (0.6)	1.2	.53 (7)
Withdrawn (20)	1.0–3.7	2.5 (0.8)	2.9	.80 (10)
Attention Problems (18)	1.6–4.8	3.5 (1.0)	3.5	.84 (9)
Aggressive Behavior (50)	3.5–10.4	6.9 (2.4)	7.7	.93 (25)
DSM–Affective Problems (16)	0.5–2.2	1.3 (0.5)	2.0	.75 (8)
DSM–Anxiety Problems (14)	0.5–2.8	1.6 (0.7)	1.9	.66 (7)
DSM–Pervasive Developmental Problems (26)	1.2–4.9	3.2 (1.1)	3.2	.76 (13)
DSM–Attention-Deficit/Hyperactivity (26)	2.8–8.1	5.5 (1.7)	5.0	.88 (13)
DSM–Oppositional Defiant Problems (14)	1.1–3.0	2.0 (0.6)	2.5	.82 (7)
Stress Problems (14)	0.9–2.5	1.5 (0.4)	1.8	.63 (7)

Note: C-TRF = *Caregiver–Teacher Report Form* (Achenbach & Rescorla, 2000); DSM = *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; American Psychiatric Association, 1994).

^aRange for society means.

^bAverage of 15 societal means and 15 societal standard deviations.

^cOmnicultural mean of 15 societal alphas; number of items per scale in parentheses.

effects could be statistically significant, we used a probability value of .001.

To achieve the first purpose of the study, we tested the effects of society, gender, and age on 15 problem scales. Because three societies had few or no children younger than 4 years, age in years was used as a continuous covariate rather than dichotomized as younger versus older. Multivariate analyses of covariance (MANCOVAs) were used when multiple scales with nonoverlapping items could be tested in a single analysis (i.e., the six syndromes, the five DSM-oriented scales, and Internalizing/Externalizing), whereas separate ANCOVAs were used for Total Problems and Stress Problems. ESs were measured by partial eta squared (η^2), which represents the percentage of total variance uniquely accounted for by a given variable with the other variables partialled out. ESs were interpreted using Cohen's (1988) criteria (*small* = 1–5.9%, *medium* = 6–13.9%, and *large* \geq 14%). Student-Newman-Keuls post-hoc tests (with $p < .001$) were used to test for differences between societies.

To test similarities in item ratings across societies, mean ratings for each of the 99 items were calculated by averaging all the 0–1–2 ratings obtained on each item within each society. Correlations between these 99 mean item ratings were then computed for every pair of societies. To compare the internal consistencies of problem scales across societies, Cronbach's alphas were computed for each scale within each society, and correlations were then computed between alphas for every pair of societies.

To achieve the second major purpose of the study, we tested societal effects on cross-informant agreement for the 7,380 preschoolers in the 13 societies with ratings by a caregiver/teacher and a parent. The 17 items differing between the C-TRF and the CBCL/1½–5 were excluded, yielding 82 items in common for cross-informant analyses. Raw scale scores were calculated for both forms based on the common 82-item set. Because 3 societies had no or very few children younger than age 4 and preliminary analyses suggested that cross-informant patterns varied somewhat by age, we conducted repeated-measures ANOVAs on Total Problems, Internalizing, and Externalizing separately by age group (1½–3, 4–5) to test for effects of informant (parents vs. caregivers/teachers, repeated measures), society, and gender across the 13 societies. Correlations were also computed between CBCL/1½–5 and C-TRF problem scale scores within each society. Finally, we tested agreement between parents and caregivers/teachers on item ratings, first using mean item ratings from caregivers/teachers within each society and then using parent and caregiver/teacher item ratings separately for each child.

Results

Societal, Gender, and Age Effects on Mean Scale Scores

For each scale, Table 2 displays the range of mean scores, the omnicultural mean (derived by averaging the 15 society means; Ellis & Kimmel, 1992) and its standard deviation,

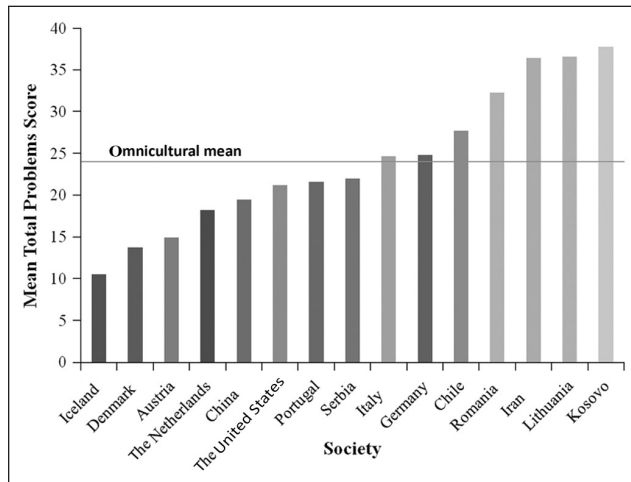


Figure 1. Mean C-TRF Total Problems scores in 15 societies ($N = 10,521$)

Note: C-TRF = Caregiver-Teacher Report Form. The omnicultural mean = 24.1 ($SD = 8.6$), with a possible range of 0 to 198.

as well as the omnicultural standard deviation (derived by averaging the 15 SD s). The omnicultural mean for Total Problems was 24.1 ($SD = 8.6$). Mean Total Problems scores in ascending order for the 15 societies are displayed in Figure 1. On a scale that could range from 0 to 198, all 15 societies had scores between 10.6 and 37.8, with 9 of the 15 societies scoring within 8.6 points (1 SD) of the omnicultural mean. In all, 3 societies (Iceland, Denmark, and Austria) had scores greater than 1 SD below the omnicultural mean, whereas 3 other societies (Kosovo, Lithuania, and Iran) had scores greater than 1 SD above the omnicultural mean. The omnicultural standard deviation (the mean of the 15 SD s) was 20.2, more than double the standard deviation of the omnicultural mean (8.6), showing that Total Problems scores varied much more within than between societies.

Table 3 displays the ESs for significant effects of society, gender, age, and the society \times gender Interactions for all scales. No other interactions were significant. Each ES indicates the percentage of variance in scores on a particular scale accounted for by a particular factor with the other factors partialled out. For example, entries in the first row of Table 3 indicate the percentage of variance in Total Problems scores accounted for by society, gender, age, and society \times gender interactions. F and p values are not included in the table due to the large number of effects and the fact that only ESs with $p < .001$ were considered significant. All multivariate effects in the MANCOVAs were significant at $p < .001$, but for clarity of presentation these are not presented in Table 1.

The ANCOVA for Total Problems yielded a medium ES (12%) for differences between societies. Gender yielded a 1% ES, indicating a slight tendency for boys to score higher

than girls on Total Problems. The age covariate was significant ($ES < 1\%$), indicating a slight tendency for younger children to score higher than older children. The society \times gender interaction was not significant. S-N-K post hoc tests indicated that no society had a mean Total Problems score significantly higher or significantly lower than all other societies.

As can be seen in Table 3, the ESs for society were 12% for Internalizing and 7% for Externalizing. Boys scored significantly higher than girls, with a smaller gender effect for Internalizing ($ES < 1\%$) than for Externalizing ($ES = 3\%$). The significant but very small society \times gender interaction for Externalizing ($< 1\%$) indicated that the magnitude of the gender effect varied somewhat across societies. Younger preschoolers had higher scores than older preschoolers on Internalizing and Externalizing ($ES < 1\%$ for both). For both Internalizing and Externalizing, Iceland, Denmark, and Austria had the lowest mean scores, but they were not significantly different from each other nor from the other societies. For Internalizing, Kosovo, Lithuania, and Iran had the highest mean scores, but they were not significantly different from each other or from those of the other societies. For Externalizing, Kosovo, Lithuania, Iran, Romania, and Chile had significantly higher scores than all the other societies.

As shown in Table 3, the MANCOVAs for the two sets of narrow-band scales (syndrome scales and *DSM*-oriented scales) yielded ESs for differences between societies ranging from 5% (*DSM*-Affective Problems, *DSM*-Oppositional Defiant Problems) to 15% (Somatic Complaints). Boys scored significantly higher than girls on nine narrow-band scales, with all ESs less than or equal to 1% except for Attention Problems (3%), Aggressive Behavior (2%), and *DSM*-Attention-Deficit/Hyperactivity Problems (2%). The society \times gender interaction was significant for Attention Problems and *DSM*-Attention-Deficit/Hyperactivity Problems ($ES < 1\%$), indicating that the size of the gender effect varied slightly across societies. Girls scored significantly higher than boys only on Somatic Complaints ($ES < 1\%$). Younger preschoolers scored significantly higher than older preschoolers on all narrow-band scales except Withdrawn and *DSM*-Affective Problems, but all ESs were very small ($< 1\%$). The ANOVA for Stress Problems yielded an ES for Society of 6%, with boys scoring slightly higher than girls ($ES < 1\%$).

Although Iceland, Denmark, and Austria had the lowest mean scores on most narrow-band scales, their mean scores were significantly lower than those of all other societies only on Attention Problems and *DSM*-Attention-Deficit/Hyperactivity Problems. Lithuania, Iran, and Kosovo had the highest mean scores on most narrow-band scales, with significantly higher scores than all other societies on the *DSM*-Pervasive Developmental Problems scale. On Somatic Complaints, Kosovo had a significantly higher mean score

Table 3. Significant Effect Sizes (η^2) for C-TRF Scale Scores

C-TRF Scale	Society (%)	Gender (%)	Age (%)	S × G (%)
Total Problems	12 ^a	1	<1	<1
Internalizing	12	<1	<1	ns
Externalizing	7	3	<1	<1
Emotionally Reactive	9	<1	<1	ns
Anxious/Depressed	9	ns	<1	ns
Somatic Complaints	15	<1	<1	ns
Withdrawn	6	<1	ns	ns
Attention Problems	6	3	<1	<1
Aggressive Behavior	7	2	<1	ns
DSM–Affective Problems	5	<1	ns	ns
DSM–Anxiety Problems	9	ns	<1	ns
DSM–Pervasive Developmental Problems	7	<1	<1	ns
DSM–Attention-Deficit/Hyperactivity Problems	6	2	<1	<1
DSM–Oppositional Defiant Problems	5	1	<1	ns
Stress Problems	6	<1	ns	ns

Note: C-TRF = Caregiver–Teacher Report Form (Achenbach & Rescorla, 2000); DSM = Diagnostic and Statistical Manual of Mental Disorders (4th ed.; American Psychiatric Association, 1994); S = Society; G = gender.

^aEffect sizes are percentage of variance accounted for by each effect that was significant at $p < .001$. No other effects were significant.

than Iran, which had a significantly higher mean score than all other societies. On the rest of the scales, the lowest and highest scoring societies were not significantly different from the remaining societies. Overall, results suggested that the societies with lower scores (or higher scores) showed the same tendency on all problem scales, rather than only on a particular type of problem scale. Similarly, societies scoring within 1 standard deviation of the omnicultural mean tended to do so on all problem scales.

Societal Effects on Mean Item Ratings

For each society, we averaged the 0–1–2 ratings of each item for the entire sample from that society. We computed Pearson's r s between the 99 mean item ratings from each society with those from each other society, yielding a matrix of bisociety correlations. These were technically Q correlations because they were calculated over items between two "cases" (i.e., 2 societies), rather than over cases between two items. However, for simplicity here and elsewhere, we refer to Q correlations as r s. Bisociety r s between mean item ratings for all pairs of societies ranged from .49 (Austria with Kosovo) to .91 (Germany with Austria). When we averaged the bisociety r s for each society, the mean bisociety r s ranged from .64 (Kosovo) to .79 (the United States), with 13 of the 15 societies having a mean $r \geq .70$. The omnicultural mean of these 15 bisociety r s was .73, indicating strong similarity with regard to which C-TRF items received relatively high versus relatively low mean ratings across the 15 societies.

To further explore multicultural findings at the item level, we used the following procedure to identify the 10

problems that had the highest mean ratings. For each item, we averaged the mean item rating from each of the 15 societies to obtain 99 omnicultural mean item ratings. The 10 problems with the highest omnicultural mean item ratings comprised the "Top 10" problem list (see Table 4). Table 4 also lists the number of societies for which these problems made each society's own Top 10 list. When listed in descending order within each society, adjacent mean item ratings were often identical or differed only minutely. Consequently, Table 4 also indicates in parentheses the number of societies for which each problem was among the "Top 15" problems. All but 1 of the Top 10 problems was among the Top 15 problems for at least 11 societies. In all, 7 of the Top 10 problems reflect difficulties with attention, concentration, and self-control (Items 59, 5, 6, 8, 64, 96, and 16) and are on the broad-band Externalizing scale. Of the remaining 3 problems, 1 loaded on the Anxious/Depressed syndrome of the broadband Internalizing scale (Item 33) and 2 (Items 73 and 3) did not load on any syndrome. Three additional problems made the Top 10 lists for at least 5 societies when societies were analyzed separately—51 = *fidgets*; 68 = *self-conscious or easily embarrassed*; and 24 = *difficulty following directions*.

Both the bisociety correlations between mean item ratings and the Top 10 problem analysis indicated considerable consistency with respect to the rank ordering of item ratings by caregivers/teachers in 15 societies. Nevertheless, the 15 societies varied with respect to how many problems in their own Top 10 problem set matched the Top 10 set averaged across all societies (e.g., 8 problems matched for Italy, Romania, and Portugal; 7 problems matched for Iran, Germany, and Austria; 6 problems matched for Denmark,

Table 4. Ten Highest Scoring C-TRF Problems Across 15 Societies

Item no.	Item	Mean rating ^a	Societies ^b (n)	Rated 1 ^c (%)	Rated 2 ^d (%)
59	Quickly shifts from one activity to another ^e	0.54	13 (15)	34	10
33	Feelings are easily hurt ^e	0.53	12 (14)	37	8
5	Can't concentrate, can't pay attention for long	0.51	12 (14)	32	10
6	Can't sit still, restless, or hyperactive ^e	0.51	11 (14)	29	11
8	Can't stand waiting, wants everything now ^e	0.50	11 (13)	29	10
64	Inattentive, easily distracted	0.48	10 (13)	31	8
96	Wants a lot of attention ^e	0.46	5 (11)	26	9
73	Too shy or timid	0.45	10 (12)	31	7
3	Afraid to try new things	0.42	7 (8)	30	6
16	Demands must be met immediately ^e	0.42	5 (11)	27	8

Note: C-TRF = Caregiver–Teacher Report Form; For all items, 0 = not true, 1 = somewhat or sometimes true, 2 = very true or often true.

^aMean rating = omnicultural mean of 15 societal mean ratings.

^bNumber of societies for which problem was in its “Top 10” list (number of societies for which problem was in its “Top 15” items).

^cOmnicultural mean of percentage of children rated 1 across the 15 societies.

^dOmnicultural mean of percentage of children rated 2 across the 15 societies.

^eProblem was also on “Top 10” list for parents’ ratings on the *Child Behavior Checklist for Ages 1½–5* (CBCL/1½–5) in 24 societies (Rescorla et al., 2011).

the Netherlands, the United States, China, and Serbia; and 5 problems matched for Lithuania, Iceland, Chile, and Kosovo).

Also listed in Table 4 is the omnicultural mean of the percentage of children across the 15 societies whose caregivers/teachers gave ratings of 1 = *somewhat or sometimes true* or 2 = *very true or often true* to the 10 highest scoring problems. For each item, the omnicultural mean percentage was obtained by averaging the percentage of preschoolers rated 1 and the percentage rated 2 in the 15 societies. These 10 omnicultural mean percentages ranged from 28% to 39% for ratings of 1 and from 7% to 12% for ratings of 2. Problems were about 3 times more likely to be rated 1 than 2, indicating that more children manifested these problems to a mild degree than to a severe degree. As indicated in Table 4, 6 of the Top 10 problems were also among the Top 10 problems reported by parents of preschoolers across 24 societies (Rescorla et al., 2011).

Societal Effects on Internal Consistency

When Cronbach’s alphas for all 15 scales in each society were correlated with alphas for all 15 scales in every other society, bisociety correlations ranged from .85 (Portugal with Iceland) to .99 (Italy with Iceland), with a mean bisociety r of .93. Mean bisociety r s for each society ranged from .90 for the Netherlands to .95 for the United States, China, Romania, Germany, and Austria. These large correlations indicate that the internal consistencies of the C-TRF scales were very similar across societies, with Total Problems, Internalizing, Externalizing, Aggressive Behavior, and *DSM–Attention-Deficit/Hyperactivity* Problems having the largest alphas. For each scale, we

averaged the alphas across societies to yield an omnicultural mean alpha, as shown in Table 2. Omnicultural mean alphas for Total Problems, Internalizing, and Externalizing were .95, .88, and .94, respectively. As Table 2 shows, 4 narrow-band scales had mean alphas less than .70, 4 had mean alphas from .70 to .79, and 4 had mean alphas greater than .80.

Societal Effects on Cross-Informant Agreement: Mean Score Comparisons

To achieve the second major purpose of the study, we used separate repeated-measures ANOVAs of scale scores within each age group for the cross-informant sample of 7,380 from 13 societies (ages 1½–3 = 34%, ages 4–5, 66%). We tested societal effects on cross-informant agreement for Total Problems, Internalizing, and Externalizing, with the informant (parents vs. caregivers/teachers) as the within-participants factor and with society and gender as the between-participants factors. ESs from the 2 (Informant) \times 13 (Society) \times 2 (Gender) ANOVAs are reported below for all effects significant at $p < .001$.

For ages 1½ to 3, the ANOVA on Total Problems yielded significant main effects for informant (11%), society (25%), and gender (2%), and the informant \times society interaction (5%). For ages 4 to 5, effects were significant for informant (13%), society (18%), gender (1%), the informant \times society interaction (3%), and the informant \times gender interaction (1%). CBCL/1½–5 mean Total Problems scores were significantly higher than C-TRF mean Total Problems scores in all 13 societies, but the informant effect was larger in some societies (e.g., China, Iceland) than in others (e.g., Iran, Denmark). The informant effect was comparable for

Table 5. Cross-Informant Correlations Between C-TRF and CBCL/1½–5 Scales

Scale	Minimum	Maximum	Omnicultural mean ^a
Total problems	.18	.49	.29
Internalizing	.13	.47	.25
Externalizing	.21	.57	.35
Emotionally Reactive	.09	.37	.22
Anxious/Depressed	.16	.41	.24
Somatic Complaints	.06	.30	.17
Withdrawn	.12	.52	.28
Attention Problems	.21	.50	.34
Aggressive Behavior	.19	.55	.33
DSM–Affective	.04	.45	.18
DSM–Anxiety	.13	.40	.25
DSM–Pervasive Developmental	.14	.50	.32
DSM–Attention Deficit/Hyperactivity	.15	.50	.31
DSM–Oppositional Defiant	.14	.51	.29
Stress Problems	.11	.48	.25

Note: C-TRF = Caregiver–Teacher Report Form; CBCL/1½–5 = *Child Behavior Checklist for Ages 1½–5* (Achenbach & Rescorla, 2000); DSM = *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; American Psychiatric Association, 1994).

^aMean of the within-society *r*s for the 13 societies with yoked C-TRF and CBCL/1½–5 data.

younger and older boys (9% and 10%, respectively), but smaller for younger girls than older girls (13% and 18%, respectively).

For ages 1½ to 3, ANOVAs indicated that the informant effect for Internalizing (1%) was much smaller than that for Externalizing (16%). By contrast, societal effects were larger for Internalizing (27%) than for Externalizing (15%). The informant × society interaction was 5% for both scales. Gender had a significant effect for Externalizing only (3%). For Internalizing, the informant ES for boys (<1%) was comparable with that for girls (1%). However, for Externalizing, the informant ES for boys (14%) was smaller than that for girls (19%).

For ages 4 to 5, the informant effect was smaller for Internalizing (5%) than for Externalizing (15%), but societal effects were similar for the two scales (Internalizing = 16%, Externalizing = 13%). The effect of gender was only significant for Externalizing (2%). The informant × society interaction was the same for Internalizing and Externalizing (3%), as was the informant × gender interaction (<1%). For Internalizing, the informant ES was smaller for boys (3%) than for girls (7%). For Externalizing, there was a much larger gender disparity for the informant effect (boys = 10% and girls = 23%).

It is noteworthy that the repeated-measures ANOVAs on C-TRF and CBCL/1½–5 Total Problems, Internalizing, and Externalizing in the cross-informant sample ($n = 7,380$) yielded larger ESs for society than those obtained in the C-TRF analyses for the full sample ($N = 10,521$). To explore this finding, we computed separate societal ESs for C-TRF and CBCL/1½–5 Total Problems using the cross-informant

sample of 7,380 children. For ages 1½ to 5, societal ESs were 19% on the C-TRF and 18% on the CBCL/1½–5, both smaller than the societal ES of 25% in the repeated-measures analysis but larger than the societal ES of 12% for the full C-TRF sample of 10,521 from 15 societies. Similarly, for ages 4 to 5, societal ESs were 15% on the C-TRF and 13% on the CBCL/1½–5, both smaller than the societal ES of 18% in the repeated-measures ANOVA and only slightly higher than the societal ES of 12% for the full C-TRF sample of 10,521. We also computed the correlation between mean C-TRF and CBCL/1½–5 Total Problems scores for the 13 cross-informant samples, which was .87, indicating that societies with low C-TRF scores also tended to have low CBCL/1½–5 scores. It thus appears that the cross-informant, repeated-measures analyses magnified societal differences in score levels because, when a society deviated widely from the omnicultural mean on the C-TRF, it also deviated widely on the CBCL/1½–5.

Societal Effects on Cross-Informant Correlations for Scale Scores

Societal effects on cross-informant agreement were also tested by comparing cross-informant *r*s computed within each society. These cross-informant *r*s indicated the degree of association between problem scale scores derived from ratings by parents versus caregivers/teachers of the same children within each of 13 societies. Although cross-informant *r*s were significant at $p < .001$ for most problem scales in most societies, nonsignificant *r*s were found on many scales for Iceland and Portugal and on a few scales for several other societies. Table 5 presents the *r* for the society

with the lowest and the highest cross-informant r , as well as the omnicultural mean r obtained by averaging the 13 societal cross-informant r s for each scale. Omnicultural mean cross-informant r s ranged from .17 for Somatic Complaints to .35 for Externalizing.

We used the within-participants asymptotic variance z test for each society to determine whether cross-informant r s were significantly larger for Externalizing than for Internalizing, relative to the within-informant r s for the two scales. The z value was significant at $p < .001$ for the United States, Germany, Romania, Portugal, and Serbia; at $p < .01$ for Lithuania; and at $p < .05$ for Denmark and Italy, indicating that cross-informant agreement was higher for Externalizing than for Internalizing in these but not other societies.

When cross-informant r s were averaged within each society across all problem scales, they ranged from .18 in Iceland to .46 in Romania. Iceland, China, and Portugal had mean problem scale cross-informant r s $< .20$; Italy, Lithuania, the Netherlands, Germany, Austria, and Serbia had mean r s between .20 and .29; Iran, Denmark, and the United States had mean r s from .30 to .39; and Romania had a mean r of .46. To test if the societies differed significantly in mean problem scale cross-informant r , we used the between-participants Fisher's z test, starting with the most extreme values to avoid Type 1 errors arising from multiple comparisons. Results indicated that Iceland's mean cross-informant r of .18 was significantly smaller at $p < .001$ than Romania's mean cross-informant r of .46, but it was not significantly smaller than the r of .33 from the United States, the next largest r after Romania.

Societal Effects on Cross-Informant Agreement for Item Ratings

The first method used to examine societal effects on cross-informant agreement for problem items used mean item ratings for parents and caregivers/teachers within each society. First, the mean ratings for parents and caregivers/teachers in each society were computed for the 82 items used for cross-informant analyses. Then, correlations were computed across the 82 mean item ratings from parents and caregivers/teachers, yielding a cross-informant r for each society. These r s ranged from .67 in Austria to .95 in Romania (all $p < .001$), which were all large (Cohen, 1988). The omnicultural mean of the 13 r s was .80 ($SD = 0.09$). Thus, parents and caregivers/teachers within each of the 13 societies showed strong agreement, on average, with respect to which items they rated high, medium, or low.

The second method we used to examine cross-informant agreement on item ratings was to calculate within-child correlations between the parent and caregiver/teacher item ratings for each child. Mean within-child r s ranged from .17 for China to .41 for Romania, with an omnicultural mean of

.25 ($SD = 0.07$). Within-child r s were much lower than the r s based on mean item ratings for parents and caregivers/teachers within each society, indicating great within-society variation across children in their patterns of 0–1–2 ratings. We converted these r s to Fisher's z s and then submitted them to 13 (Society) \times 2 (Gender) ANOVAs, separately by age group, to test effects of society and gender on within-child item agreement between parents and caregivers/teachers. These ANOVAs yielded societal ESs of 9% for younger preschoolers and 13% for older preschoolers. Neither gender nor the Society \times Gender interaction was significant for either age group.

Discussion

The first purpose of this study was to test societal effects on caregiver/teacher ratings of 10,521 children ages 1½ to 5 from 15 societies. The second purpose was to test societal effects on cross-informant agreement between parents' and caregivers/teachers' ratings of 7,380 preschool children from 13 societies. To our knowledge, this is the first study to aggregate data from many societies to directly test societal effects on preschoolers' problems reported by caregivers/teachers and on cross-informant agreement between parents and caregivers/teachers. Overall, our findings indicated substantial similarities among societies but also some notable differences.

Of the 15 significant societal effects on mean scale scores, 2 were small (1%–5.9%), 12 were medium (5.9%–13.9%), and 1 was large (15% for Somatic Complaints). For Total Problems, the societal ES was 12%. In all, 9 of the 15 societies scored within 1 standard deviation (8.6 points) of the omnicultural mean of 24.1 on Total Problems, a scale that could range from 0 to 198. These 9 societies (Italy, the Netherlands, China, the United States, Portugal, Serbia, Germany, Chile, and Romania) differ greatly in ethnicity, religion, political/economic system, history, geographic region, and cultural values. They also most likely differ in the percentage of children enrolled in early childhood programs and in class size, curriculum, behavioral expectations, and training of staff. Nevertheless, ratings by caregivers and teachers in these societies yielded mean Total Problems scores spanning a relatively small proportion of the possible range of scores.

It is also important to underscore that 3 societies scored greater than 1 standard deviation below the omnicultural mean on Total Problems (Iceland, Denmark, and Austria) and 3 other societies scored greater than 1 standard deviation above the omnicultural mean (Lithuania, Kosovo, and Iran). It is not obvious what features the low-scoring societies (or high-scoring societies) have in common, nor how either of these groups differs systematically from the middle-scoring group. For example, Iceland and Denmark are Nordic societies, but Austria is not. Austria had lower scores than Germany, the society to which it is probably

most similar. Lithuania and Kosovo but not Iran were formerly in the “East Bloc,” but so was Serbia, which did not have particularly high scores.

The full range of C-TRF mean Total Problems scores across the 15 societies was from 10.6 to 37.8 on a scale ranging from 0 to 198. To take account of these differences between societies, multicultural norms for the C-TRF (Achenbach & Rescorla, 2010) have three norm groups (low, medium, and high), which are based on each society’s mean Total Problems score. The availability of three sets of norms based on differences in mean Total Problems scores can facilitate identification of children who may need help. For example, a Lithuanian preschooler’s Externalizing score might be in the clinical range (>97th percentile) for the medium norm group (based on U.S. norms), but in the normal range for the high norm group, which includes Lithuania. If U.S. norms are used, a clinician might assume the child requires help, but if norms appropriate for Lithuania are used, the child will not have an elevated score based on caregiver/teacher reports in Lithuania. As data become available for other societies, they will be used to determine the appropriate norm groups for those societies.

The societal ES was larger for Internalizing than for Externalizing (12% vs. 7%). Parents’ reports on the CBCL/1½–5 (Rescorla et al., 2011) also yielded a larger societal ES for Internalizing (10%) than for Externalizing (7%). Multicultural comparisons of ratings by teachers and parents of children ages 6 to 11 (Rescorla, Achenbach, Ginzburg, et al., 2007; Rescorla, Achenbach, Ivanova, et al., 2007) have likewise yielded larger societal ESs for Internalizing (11% for teachers, 9% for parents) than for Externalizing (4% for teachers, 6% for parents). This pattern of results indicates that caregivers, teachers, and parents in different societies vary more in how they rate Internalizing problems such as anxiety, depression, social withdrawal, and somatic complaints than in how they rate Externalizing problems such as aggression. This finding could reflect bigger differences across societies in the “actual” prevalence of Internalizing problems than of Externalizing problems. However, this finding might also suggest that there is more “cultural universality” in how adults view children’s Externalizing behavior than in how they view children’s Internalizing behavior, with more “cultural specificity” in judgments of problems such as anxiety, somatic complaints, and depression.

The mean bisociety r of .73 indicates that item ratings by caregivers/teachers in 15 societies were quite consistent in terms of which problem items tended to receive high, medium, or low ratings. This suggests that, despite the vicissitudes of translation, the items operated similarly in very different societies. Seven of the Top 10 problems were among the Top 10 problems for at least 10 of the 15 societies. Furthermore, 6 of the Top 10 problems were also among the Top 10 problems rated by parents (Rescorla et al., 2011), namely, 59 = *quickly shifts from one activity to another*;

33 = *feelings are easily hurt*; 8 = *can’t stand waiting, wants everything now*; 6 = *can’t sit still, restless, or hyperactive*; 16 = *demands must be met immediately*; and 96 = *wants a lot of attention*. These findings suggest that there is considerable “universality” across diverse societies with respect to the behavioral and emotional problems most commonly displayed by young children, whether viewed by their parents or their caregivers/teachers.

The mean bisociety r of .93 between alphas across the 15 societies indicates that the problem scales had similar internal consistencies across very diverse societies. In every society, the scales with the largest alphas were Total Problems, Internalizing, Externalizing, and Aggressive Behavior. These C-TRF internal consistency findings are consistent with those reported by Rescorla et al. (2011) for the CBCL/1½–5, by Rescorla, Achenbach, Ivanova, et al. (2007) for the CBCL/6–18, and by Rescorla, Achenbach, Ginzburg, et al. (2007) for the TRF. That is, Total Problems, Internalizing, Externalizing, and Aggressive Behavior, which are the scales with the most items, tended to have the largest Cronbach’s alphas across societies on all four instruments. The large bisociety r s between alphas complement Ivanova et al.’s (2011) CFA results, which showed that the data from the 14 non-U.S. samples analyzed for the present study demonstrated generally good fit to the factor model derived from U.S. samples

An important finding of our research is that mean Total Problems scores derived from ratings by caregivers/teachers were significantly lower than those derived from ratings of the same 82 items by parents in all 13 societies where preschoolers were rated by both kinds of informants. The informant disparity was smaller in some societies than in others, but it is not clear what the societies with smaller disparities (such as Denmark and Iran) have in common, nor how they differ from societies with larger disparities (such as China or Iceland). To our knowledge, only one study conducted by investigators not involved in the current study has directly compared the magnitude of parent and caregiver/teacher ratings for preschoolers, namely, the study by Winsler and Wallace (2002), which found that parents reported more problems than teachers for a sample of 47 U.S. preschoolers. There are several reasons why parents might report more problems than caregivers/teachers, including that they generally spend more time with their own children than do caregivers/teachers and thus have more opportunity to observe problems, that they see behaviors as more problematic than do caregivers/teachers because they do not see a reference group of other children in the same setting, that they may be less experienced at dealing with preschoolers than are caregivers/teachers and thus less able to forestall problems, or that they are more anxious and concerned about how their own children are doing than are caregivers/teachers who work with groups of children. However, there may be reasons for caregivers/teachers to report more problems than parents. For example, young

children might be more stressed at day care/school than at home because they are separated from parents, compelled to follow school routines, and required to adjust to other children. Despite these factors, caregiver/teacher ratings were significantly lower than parent ratings in all 13 societies with cross-informant data, suggesting that these factors were less important than the factors contributing to higher parent scores.

Another important finding is that the discrepancy between parent scores and caregiver/teacher scores was smaller for boys than for girls, especially in the older age group. This was seen in the informant ESs for the younger group (10% for boys vs. 13% for girls) compared with those for the older group (9% for boys vs. 18% for girls). These findings indicate that caregivers/teachers tend to report fewer problems than parents do for girls but less so for boys, particularly when the preschoolers are 4 to 5 years of age. This may be because, particularly as preschoolers get older, girls manifest better adjustment to the demands of the day care/school environment than do boys, with caregiver/teachers perceiving boys as being more aggressive, non-compliant, and inattentive than girls. This interpretation of our data is supported by the fact that the largest C-TRF gender ESs in the current study were on Externalizing (3%) and its associated scales (Attention Problems = 3%, Aggressive Behavior = 2%). In addition, in comparing TRF ratings from 21 societies for students ages 6 to 15, Rescorla, Achenbach, Ginzburg, et al. (2007) found that the largest gender ESs were on Externalizing, Attention Problems, Aggressive Behavior, and Rule-Breaking Behavior.

Cross-informant agreement between caregivers/teachers and parents was first tested for the 15 problem scales within each society. When cross-informant r s for all problem scales were averaged within society, mean cross-informant r s for problem scales ranged from .18 in Iceland to .46 in Romania. Iceland, Portugal, and China had the lowest agreement (.18–.19), whereas Romania, the United States, Iran, and Denmark had the highest agreement (.33–.46). However, most of the differences between societies in cross-informant r s were not significant. It is not apparent what cultural factors might account for the relatively low cross-informant agreement in Iceland, Portugal, and China. Yet, even the highest agreement was modest, consistent with Achenbach et al.'s (1987) meta-analytic findings. This is to be expected, given that caregivers/teachers interact with preschoolers in group settings, whereas parents interact with preschoolers primarily in family settings.

Cross-informant agreement between parents and caregivers/teachers on item ratings was tested two ways, first using mean item ratings from parents and caregivers/teachers within each society and then using parent and caregiver/teacher item ratings separately for each child. When r s between mean item ratings for parents and caregivers/teachers within each society were computed, r s ranged from

.67 in Austria to .95 in Romania. Thus, on average, parents and caregivers/teachers in every society agreed very well with respect to which items they tended to rate low, medium, or high. However, when “within-child” cross-informant r s for item ratings were computed, r s varied widely in every society and the mean r ranged from only .17 for China to .41 for Romania, indicating great within-society variation in agreement between parent versus caregiver/teacher ratings of particular problem items. In every society, some parent–caregiver/teacher pairs agreed well on the child's problems, whereas others agreed poorly. Societies differed in their overall level of within-child agreement, but no society had a mean r that would be considered “large” range according to Cohen's (1988) criteria.

Limitations

For some societies, such as the United States, we believe that the sample was representative of the population with respect to region, ethnicity, and socioeconomic status (SES) of children attending day care/preschool. For other societies, such as China, the sample was large and somewhat diverse in SES, but it was recruited in only one region of a very large country (a metropolitan area between Shanghai and Nanjing) and probably did not reflect ethnic minority groups who live in other regions of China or the full SES spectrum of Chinese children attending day care/preschool. Furthermore, because we lacked SES and ethnicity information in most samples, we could not test effects of these factors on problem scores. In addition, we did not have information for each society regarding whether clinically referred children were included, although the percentage of referred children was probably small. It is also likely that societies varied to some degree in access to day care/preschool as a function of SES, with children from lower SES families overrepresented in some societies and underrepresented in other societies. We also know that sampling methods varied somewhat across societies (e.g., through schools vs. through households or birth cohorts). In summary, although we had large samples from 15 societies, indigenous investigators used regional rather than national samples in 10 societies, used somewhat different methods of obtaining the data, and could not verify how representative their samples were with respect to SES, ethnicity, and other factors. These factors limit the generalizability of our findings with respect to the population of each society.

European societies were overrepresented, although we had five non-European samples (Australia, China, Chile, Iran, the United States). Because the data were collected by indigenous investigators for various purposes, rather than being collected according to a single international sampling plan, we did not have equal representation from the whole world. However, our data provide a good benchmark

against which data from additional societies can be compared in future research.

Another possible limitation of our study is that, despite the process of translation and back translation, we cannot be certain that the problem items held identical meanings for all caregivers/teachers in every society. This is because we used an *etic* method, whereby the same standardized assessment instrument was used in 15 societies, rather than an *emic* method (Pike, 1967), whereby meanings of constructs are examined within each society. Using an *emic* method, one might explore how caregivers/teachers in the different societies interpreted the problem items, which might suggest reasons why agreement between parents and caregivers/teachers was lower in some societies than in others. Despite the fact that it was not possible to conduct interviews with caregivers/teachers in each society regarding the meaning of the C-TRF items, the large bisociety *r*s obtained for mean item ratings suggest that caregivers/teachers in different societies interpreted the items similarly. Furthermore, the finding that 6 of the “Top 10” items in caregiver/teacher ratings were also among the “Top 10” items in parent ratings suggests that parents and caregivers/teachers were interpreting the items in the same way.

Conclusions

There are many differences among the 15 societies in this study with respect to language, ethnicity, region, size, religion, political/economic system, and early childhood programs. Nevertheless, many similarities were found in mean scale scores, age and gender effects, mean item ratings, scale internal consistencies derived from caregiver/teacher ratings, and cross-informant patterns for ratings by caregivers/teachers and parents.

Authors' Notes

Author affiliations by institution: Leslie A. Rescorla, PhD (Bryn Mawr College); Thomas M. Achenbach, PhD, and Masha Y. Ivanova, PhD (University of Vermont); Niels Bilenberg, MD, PhD, Solvejg Kristensen, MHSc (University of Southern Denmark); Gudrun Bjarnadottir, PhD (Glaesibaer & Arbaer Health Clinics, Iceland); Silvia Denner, PhD (Dortmund University, Germany); Pedro Dias, PhD, and Bárbara César Machado, PhD (Catholic University of Portugal); Anca Doborean, PhD (Babes-Bolyai University, Romania); Manfred Döpfner, PhD, and Julia Plücker, PhD (University of Cologne, Germany); Alessandra Frigerio, PhD, and Rosario Montirosso, MSc (Scientific Institute “E. Medea,” Bosisio Parini [LC], Italy); Miguel Gonçalves, PhD (University of Minho, Portugal); Halldór Guðmundsson, MA (University of Iceland); Roma Jusiene, PhD (Vilnius University, Lithuania); Felipe Lecannelier, MA, Sofia P. Löbel, Paola A. Mas, Jorge T. Rodriguez, and Pamela O. Rojas (University of Desarrollo, Chile); Patrick W. L. Leung, PhD

(Chinese University of Hong Kong); Jianghong Liu, PhD (University of Pennsylvania); Jasminka Markovic (Clinical Center of Vojvodina, Serbia); Elaheh Mohammad Esmaeili, PhD (Tehran Research Institute for Education, Iran); Adelina Ahmeti Pronaj and Mimoza Shahini, MD (University Clinical Center of Kosovo); Klaus Schmeck (Psychiatric University Hospitals-Basel, Switzerland); Jaime R. Silva, PhD (Universidad de la Frontera, Chile); Jan van der Ende, MA, and Frank C. Verhulst, MD, PhD (Erasmus University Medical Center-Sophia Children's Hospital, The Netherlands).

Declaration of Conflicting Interests

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

Funding

The authors disclosed receipt of the following financial support for the research and/or authorship of this article: The C-TRF is published by the Research Center for Children, Youth, and Families at the University of Vermont, from which the first three authors receive remuneration.

References

- Achenbach, T. M., McConaughy, S. H., & Howell, C. (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. (2000). *Manual for the ASEBA preschool forms & profiles*. Burlington: University of Vermont, Research Center for Children, Youth, and Families.
- Achenbach, T. M., & Rescorla, L. (2010). *Multicultural supplement to the Manual for the ASEBA preschool forms and 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.
- Carter, A. (2010). The field of toddler/preschool mental health has arrived—On a global scale. *Journal of the American Academy of Child & Adolescent Psychiatry, 49*, 1181–1182. doi:10.1016/j.jaac.2010.09.006
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Academic Press.
- Crane, J., Mincic, M. S., & Winsler, A. (2011). Parent-teacher agreement and reliability on the Devereux Early Childhood Assessment (DECA) in English and Spanish for ethnically diverse children living in poverty. *Early Education and Development, 22*, 520–547. doi:10.1080/10409289.2011.565722
- 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

- Denner, S., & Schmeck, K. (2005). Emotional and behavioural disorders at preschool age—Results of a study in kindergarten children in Dortmund using the Caregiver–Teacher Report Form C-TRF/1½–5. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, *33*, 307–317.
- Dias, P., Machado, B. C., Silva, J., & Goncalves, M. (2009). *Child Behavior Checklist/1½–5 and C-TRF scores for Portuguese preschool children*. Unpublished manuscript.
- Dobrea, A. (2009). *Child Behavior Checklist/1½–5 and C-TRF scores for Romanian preschool children*. Unpublished manuscript.
- Döpfner, M., & Plück, J. (2009). *Child Behavior Checklist/1½–5 and C-TRF scores for German preschool children*. Unpublished manuscript.
- 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.
- Eyberg, S., & Pincus, D. (1999). *Eyberg Child Behavior Inventory and Sutter Eyberg Student Behavior Inventory—Revised*. Odessa, FL: Psychological Assessment Resources.
- Frigerio, A., Cozzi, P., Pastore, V., Molteni, M., Borgatti, R., & Montirosso, R. (2006). The evaluation of behavioral and emotional problems in a sample of Italian preschoolers using the Child Behavior Checklist and the Caregiver-Teacher Report Form. *Infanzia e Adolescenza*, *5*, 24–32.
- Gross, D., Fogg, L., Garvey, C., & Julion, W. (2004). Behavior problems in young children: An analysis of cross-informant agreements and disagreements. *Research in Nursing & Health*, *27*, 413–425.
- Guðmundsson, H., & Bjarnadóttir, G. (2009). *Child Behavior Checklist/1½–5 and C-TRF scores for Icelandic preschool children*. Unpublished manuscript.
- Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Bilenberg, N., Kristensen, S., Bjarnadóttir, G., ... Verhulst, F. C. (2011). Syndromes of preschool psychopathology reported by teachers and caregivers in 14 societies. *Journal of Early Childhood and Infant Psychology*, *7*, 99–116.
- Jusienė, R., Raižienė, S., Barkauskienė, R., Bielauskaitė, R., & Dervinytė-Bongarzoni, A. (2007). The risk factors of emotional and behavioral problems in preschool age. *Visuomenės Sveikata*, *4*, 46–54.
- Kirk, R. E. (1995). *Experimental design: Procedures for behavioral sciences* (3rd ed.). New York, NY: Wadsworth.
- Kristensen, S., Henriksen, T. B., & Bilenberg, N. (2010). The Child Behavior Checklist for Ages 1.5-5 (CBCL/1½–5): Assessment and analysis of parent- and caregiver-reported problems in a population-based sample of Danish preschool children. *Nordic Journal of Psychiatry*, *64*, 203–209. doi:10.3109/08039480903456595
- LaFreniere, P. J., & Dumas, J. E. (1995). *Social Competence and Behavior Evaluation—Preschool edition (SCBE)*. Los Angeles, CA: Western Psychological Services.
- LaFreniere, P. J., Masataka, N., Butovskaya, M., Chen, Q., Dessen, M. A., Atwanger, K., . . . Frigerio, A. (2002). Cross-cultural analysis of social competence and behavior problems in preschoolers. *Early Education and Development*, *13*, 201–219. doi:10.1207/s15566935eed13026
- LeBuffe, P. A., & Naglieri, J. A. (1999). *The Devereux Early Childhood Assessment*. Lewisville, NC: Kaplan Press.
- Lecanneler, F., Löbel, S. P., Mas, P. O., Rodriguez, J. T., & Rojas, P. O. (2009). *C-TRF scores for Chilean preschool children*. Unpublished manuscript.
- Liu, J., Cheng, H., & Leung, P. W. L. (2010). The application of the preschool Child Behavior Checklist and Caregiver–Teacher Report Form to Mainland Chinese children: Syndrome structure, gender differences, country effects, and inter-informant agreement. *Journal of Abnormal Child Psychology*, *39*, 251–264. doi:10.1007/s10802-010-9452-8
- Markovic, J. (2011). *Child Behavior Checklist/1½–5 and C-TRF scores for Serbian preschool children*. Unpublished manuscript.
- Merrell, K. W. (1994). *Preschool and Kindergarten Behavior scales: Test manual*. Brandon, VT: Clinical Psychology Publishers.
- Mohammad Esmali, E. (2009). *Child Behavior Checklist/1½–5 and C-TRF scores for Iranian preschool children*. Unpublished manuscript.
- Pike, K. L. (Ed.). (1967). *Language in relation to a unified theory of structure of human behavior* (2nd ed.). The Hague, the Netherlands: Mouton.
- Rescorla, L., Achenbach, T. M., Ginzburg, S., Ivanova, M. Y., Dumenci, L., Almqvist, F., . . . Verhulst, F. (2007). Consistency of teacher-reported problems for students in 21 countries. *School Psychology Review*, *36*, 91–110.
- Rescorla, L., Achenbach, T. M., Ivanova, M. Y., Dumenci, L., Almqvist, F., Bilenberg, N., . . . Verhulst, F. C. (2007). Behavioral and emotional problems reported by parents of children ages 6 to 16 in 31 Societies. *Journal of Emotional and Behavioral Disorders*, *15*, 130–142.
- Rescorla, L., Achenbach, T. M., Ivanova, M. Y., Harder, V. S., Otten, L., Bilenberg, N., . . . Verhulst, F. C. (2011). International comparisons of behavioral and emotional problems in preschool children: Parents' reports from 24 societies. *Journal of Clinical Child and Adolescent Psychology*, *40*, 456–467.
- Schmeck, K., & Skrabel, C. (2004). *Child Behavior Checklist/1½–5 and C-TRF scores for Austrian preschool children*. Unpublished manuscript.
- Shahini, M., & Pranvera, J. (2010). *C-TRF scores for Kosovar preschool children*. Unpublished manuscript.
- Tick, N. T., van der Ende, J., Koot, H. M., & Verhulst, F. C. (2007). 14-year changes in emotional and behavioral problems of very young Dutch children. *Journal of the American Academy of Child & Adolescent Psychiatry*, *46*, 1333–1340. doi:10.1097/chi.0b013e3181337532
- Winsler, A., & Wallace, G. L. (2002). Behavior problems and social skills in preschool children: Parent–teacher agreement and relations with classroom observations. *Early Education and Development*, *13*, 41–58. doi:10.1207/s15566935eed1301_3