School Psychology Review,
2007, Volume 36, No. 1, pp. 91–110

Consistency of Teacher-Reported Problems for Students in 21 Countries

Leslie A. Rescorla,[1] Thomas M. Achenbach,[2] Sofia Ginzburg,[1] Masha Ivanova,[2] Levent Dumenci,[2] Fredrik Almqvist,[3] Marie Bathiche,[4] Niels Bilenberg,[5] Hector Bird,[6] Anca Domuta,[7] Nese Erol,[8] Eric Fombonne,[4] Antonio Fonseca,[9] Alessandra Frigerio,[10] Yasuko Kanbayashi,[11] Michael C. Lambert,[12] Xianchen Liu,[13] Patrick Leung,[14] Asghar Minaei,[15] Alexandra Roussos,[16] Zeynep Simsek,[17] Sheila Weintraub,[3] John Weisz,[18] Tomasz Wolanczyk,[19] Stephen R. Zubrick,[20] Rita Zukauskiene,[21] Frank Verhulst[22]

[1]Bryn Mawr College, [2]University of Vermont, [3]Helsinki University, [4]McGill University, [5]University of Southern Denmark, [6]Columbia University, [7]Babes-Bolyai University, [8]Ankara University, [9]University of Coimbra, [10]Istituto Scientifico E. Medea, [11]Chuo University, [12]University of Missouri, Columbia, [13]University of Pittsburgh, [14]Chinese University of Hong Kong, [15]Research Institute of Exceptional Children, Tehran, [16]Attiki Child Psychiatric Hospital, [17]University of Harran, [18]Harvard University, [19]Medical University of Warsaw, [20]Curtin University of Technology and the Telethon Institute for Child Health Research, [21]Mykolas Romeris University, [22]Erasmus University Medical Center—Sophia Children's Hospital

Abstract. This study compared teachers' ratings of behavioral and emotional problems on the Teacher's Report Form for general population samples in 21 countries (N = 30,957). Correlations between internal consistency coefficients in different countries averaged .90. Effects of country on scale scores ranged from 3% to 13%. Gender effects ranged from <1% to 5%, and age effects were all <1%. With great consistency across countries, scores were higher for boys than for girls on eight scales: Total Problems; Externalizing; the Attention Problems, Rule-Breaking Behavior, and Aggressive Behavior syndromes; and *Diagnostic and Statistical Manual (DSM)*-oriented Attention Deficit Hyperactivity Problems, Oppositional Defiant Problems, and Conduct Problems. Correlations between mean item ratings in different countries averaged .74. Teacher's Report Form results were thus similar across 21 very diverse countries, despite differences across these countries in school systems, models of pedagogy, and curricula.

In their classic article, Cronbach and Meehl (1955) stated that construct validity addresses the question, "What constructs account for variation in test performance?" Their first example for establishing construct validity involves investigating whether a test is

We thank Stephanie H. McConaughy for her helpful suggestions on the manuscript.

Correspondence regarding this article should be addressed to Leslie Rescorla, Bryn Mawr College, 101 N. Merion Avenue, Bryn Mawr, PA 19010; e-mail: lrescorl@brynmawr.edu

"culture-free." Thus, a crucial aspect of construct validation is determining if culture accounts for systematic variation in test performance. As described by Bracken and Barona (1991) and Geisinger (1994), many steps are needed to demonstrate what might be called the *multicultural robustness* of an instrument. First, translation of the instrument may be required, which in itself is a multistep process. Next, data must be collected using a large, representative sample. Among the features of the instrument that need to be systematically compared using data from different cultures are test–retest reliability, factor structure, internal consistency of scales, performance of individual items, mean scores, associations of scores with age and sex, and correlations with validating criteria.

The present study investigated the internal consistency reliability, mean scale scores, age and gender patterns, and mean item scores of the Teacher's Report Form (TRF; Achenbach & Rescorla, 2001) for 30,957 children in 21 different societies (19 countries, plus Hong Kong and the commonwealth of Puerto Rico, hereafter all called "countries"). First published by Achenbach and Edelbrock (1986) and revised by Achenbach and Rescorla (2001), the TRF has been translated into more than 20 languages. Although more than 200 published studies have reported TRF data from countries other than the United States, no studies to date have compared TRF scores from many countries in the same analyses. Findings from the present study are complemented by confirmatory factor analyses conducted by Ivanova et al. (2006) using the same data. Ivanova et al. (2006) reported good fit for the U.S.-derived factor model in the 20 other countries. The confirmatory factor analysis findings demonstrated that the constructs tapped by TRF syndromes are consistent across countries, which is an important component of multicultural robustness.

Overview of the TRF

Along with the Child Behavior Checklist (CBCL) completed by parents and the Youth Self-Report (YSR), the TRF is a component of the Achenbach System of Empirically Based Assessment (ASEBA; Achenbach & Rescorla, 2001). The ASEBA uses a "bottom-up" approach, whereby syndromes of co-occurring problems are identified though exploratory and confirmatory factor analysis. The scales for scoring the syndromes are then normed by age and gender using data from representative population-based samples.

The 2001 edition of the TRF, designed for ages 6–18 (Achenbach & Rescorla, 2001), has 118 specific problem items, plus 2 open-ended problem items, all of which are rated as 0 = *not true (as far as you know)*; 1 = *somewhat or sometimes true;* and 2 = *very true or often true.* Teachers rate problems based on students' functioning over the preceding 2 months. Although not analyzed for this study, the TRF also assesses academic performance and adaptive characteristics.

The TRF yields problem scores at four levels of a scoring hierarchy, with the 120 problem items comprising the lowest level and the eight syndromes derived by factor analysis comprising the second level. The eight 2001 syndromes (Achenbach & Rescorla, 2001) are very similar to the TRF syndromes published by Achenbach in 1991, although different factor-analytic procedures and a new sample were used. The third level of the hierarchy contains *Internalizing* and *Externalizing* scales, derived from second-order factor analyses of the eight syndromes and obtained for both the pre-2001 and 2001 versions of the TRF. The *Anxious/Depressed, Withdrawn/Depressed* (formerly called *Withdrawn*), and *Somatic Complaints* syndromes comprise the Internalizing scale, whereas the *Rule-Breaking Behavior* (formerly called *Delinquent Behavior*) and *Aggressive Behavior* syndromes comprise the Externalizing scale. The *Social Problems, Thought Problems,* and *Attention Problems* syndromes did not load differentially on either Internalizing or Externalizing. The eight syndromes and the Internalizing and Externalizing scales are scored by summing their constituent items. The *Total Problems* scale, at the top of the hierarchy, is the sum of the ratings on all problem items.

Although not part of the empirically based scoring hierarchy, the TRF also has six *Diagnostic and Statistical Manual* (DSM)-oriented scales: *Affective Problems, Anxiety Problems, Somatic Problems, Attention Deficit Hyperactivity Problems, Oppositional Defiant Problems,* and *Conduct Problems*. These scales comprise items identified by an international panel of experts as being very consistent with diagnostic categories of the American Psychiatric Association's (1994) *Diagnostic and Statistical Manual of Mental Disorders—4th edition* (DSM-IV; Achenbach, Dumenci, & Rescorla, 2003). The TRF also has two sets of subscales tapping the Inattentive and Hyperactive-Impulsive subtypes of DSM-IV Attention Deficit Hyperactivity Disorder, one set derived from factor analysis and the other set identified by experts' ratings (Achenbach & Rescorla, 2001; Dumenci, McConaughy, & Achenbach, 2004).

Achenbach and Rescorla (2001) reported reliability and validity data for U.S. samples. Internal consistency reliability (alpha coefficients) ranged from .72 to .95 for syndromes, from .73 to .94 for DSM-oriented scales, and from .90 to .97 for Internalizing, Externalizing, and Total Problems. The mean test–retest reliability was .90 across all TRF empirically based scales and .85 across all DSM-oriented scales. Multiple regression analyses of TRF scores for demographically matched nonreferred and clinically referred samples yielded significant effects for referral status on all TRF scales, with effect sizes (ESs), as computed using eta-squared (η^2), ranging from 3% for Somatic Complaints and DSM-oriented Somatic Problems to 26% for Total Problems. Higher problem scores were significantly associated with lower socioeconomic status (SES) for most TRF scales, but the ESs were all small ($\leq$2%).

International Research Using the TRF

TRF findings for general population samples have been published from nine countries other than the United States, as summarized in the following paragraphs. In eight of these countries, the first step was to translate the TRF. Consistent with the guidelines outlined by Bracken and Barona (1991) and Geisinger (1994), the typical procedure was that a linguist or professional translator translated the TRF. An independent translator then performed a back-translation into English. The original English and the back-translated versions were reviewed by the linguists and researchers to revise ambiguous or misleading items. No translation was needed for Australia or Jamaica, where English is the official language.

The next step in each study was to recruit a large general population sample. This was done either by selecting a stratified, representative sample of households (based on community household registers or home addresses) and then contacting the children's teachers, or by selecting a stratified, representative sample of schools and then randomly sampling children within classrooms. Although analyses varied, most studies reported the effects of age, gender, and SES on problem scores.

Five studies compared teacher-reported problems for a general population sample in one country with teacher-reported problems for a U.S. sample via analysis of variance (ANOVA). ESs were quantified using η^2 and interpreted according to Cohen's (1988) criteria for the percent of variance accounted for: *small* = 1–5.9%, *medium* = 5.9–13.8%, *large* >13.8%.

The earliest of these *bicountry* studies compared teacher-reported problems for 748 Dutch and 600 U.S. 6- to 11-year-old students (Achenbach, Verhulst, Edelbrock, Baron, & Akkerhuis, 1987). Dutch and U.S. TRF Total Problems scores did not differ significantly. Weisz et al. (1989) compared teacher-reported problems for 359 Thai and 586 U.S. 6- to 11-year-old students. Thai teachers reported significantly more problems than American teachers on Total Problems (ES = 5%), Internalizing (called *overcontrolled* in that study, ES = 9%), and Externalizing (called *undercontrolled* in that study, ES = 3%).

Achenbach et al. (1990) compared teacher-reported problems for 567 Puerto Rican 6- to 16-year-old students and 567 U.S.

mainland 6- to 16-year old students matched on age, gender, and SES. Puerto Rican students obtained higher Total Problems scores than U.S. mainland students (ES = 3%). Lambert, Knight, Taylor, and Achenbach (1995) compared TRF scores for 359 Jamaican and 665 U.S. 6- to 11-year-olds. Jamaican students obtained slightly higher scores than U.S. students on Total Problems, Internalizing, Externalizing, and six syndromes (ESs $\leq$ 2%). However, Lambert, Lyubansky, and Achenbach (1998) found no significant effects of country on Total Problems, Internalizing, or Externalizing scores for 12- to 18-year-old Jamaican students. Weine, Phillips, and Achenbach (1995) compared teacher-reported problems for 469 Chinese and 469 U.S. 6- to 13-year-old students, matched across country by age, gender, and SES (based on parent education in China and parent occupation in the United States). Chinese students generally had higher TRF scores than U.S. students (ESs 2–7%), but rural American teachers rated students higher on Somatic Complaints and Aggressive Behavior than rural Chinese teachers.

All five bicountry studies reported gender differences on TRF scores, with boys typically obtaining higher scores than girls on Total Problems and Externalizing. Girls obtained higher Internalizing scores than boys among Jamaican 12- to 18-year-olds, whereas boys obtained higher Internalizing scores than girls in Thailand. Higher problem scores were reported for students from lower SES families in the Netherlands, Puerto Rico, Jamaica, and Thailand. Test–retest reliability was .84 in the Netherlands and .91 in Thailand and Jamaica. Correlations between mean ratings on all TRF items were .78 between Puerto Rican and U.S. mainland samples, .72 between Chinese and U.S. urban samples, and .44 between Chinese and U.S. rural samples.

Five additional studies examined teacher-reported problems in a single country, but no statistical comparisons were made to U.S. scores. For Greek students (N = 1,179; Roussos et al., 1999) and Portuguese students (N = 1,586; Fonseca et al., 1995), mean Total Problems scores were reported to be close to U.S. scores. Portuguese boys had higher scores than girls on Externalizing scales, and younger students had higher Total Problems scores than older students. Frigerio et al. (2004) reported TRF data for 1,464 Italian students. Boys scored significantly higher than girls on most scales except the three Internalizing syndromes, and younger students scored significantly higher than older students on most scales. Bilenberg (1999) reported TRF data for 547 Danish students. TRF scores were significantly higher for boys than for girls. In Greece, Italy, and Denmark, lower SES was associated with higher TRF scores. Finally, Liu et al. (2000b) reported TRF scores for a sample of 2,936 children 6–11 years old in Shandong, China. Boys obtained higher scores than girls on Total Problems, Externalizing, and several syndromes (ES <1%). Test–retest reliability was .83 for Total Problems.

Achenbach and Rescorla (2007) have reviewed findings from these studies and from other international studies that tested associations of TRF scores with either psychiatric diagnoses or clinical referral status. Significant associations between TRF scores and psychiatric diagnoses or referral status have been reported for the Netherlands, Puerto Rico, and Denmark, as well as for the United States, thus supporting criterion-related validity of the TRF in multiple countries.

Goals of the Present Study

The present study systematically compared TRF scores from 21 countries to evaluate similarities and differences in how teachers around the world rate their students' behavioral and emotional problems. Some findings from the Puerto Rican, Thai, Jamaican, Portuguese, Greek, Italian, Danish, and Chinese samples have been published previously, as summarized earlier in this article. However, the present study extended previous research by comparing TRF scores from 21 countries using uniform analytic procedures. Because these countries differ widely in language, geographic region, political and economic system, racial and ethnic composition, religion, and educational system, comparing them provides

a comprehensive test of the generalizability of problems reported by teachers on the TRF.

The present study addressed the following five major questions: (a) How did the 21 countries compare with respect to internal consistency reliabilities for TRF problem scales? (b) How large were the effects of country on TRF scales? (c) How consistent were effects of gender and age on TRF scores across the 21 countries? (d) To what degree did teachers in different countries tend to give high, medium, or low ratings to the same problem items? (e) Did differences between countries in per capita income, the number of students rated by each teacher, response rates, or inclusion versus exclusion of students referred for mental health or special education services affect TRF Total Problems scores?

Method

Samples

For samples from 19 countries plus Hong Kong and Puerto Rico, Table 1 lists the reference for each data set, plus the following characteristics: total *N*, ages analyzed, sampling procedure, per capita income ranking of the country, number of TRFs rated by each teacher, response rate, and whether referred students were excluded. In 8 of the countries, children were selected in household surveys; parents then consented to having their children's teacher complete the TRF. In 13 countries, schools provided the sampling frame.

SES information was lacking for students in many samples, so the association between SES and TRF scores within countries could not be tested. However, based on World Bank (2005) statistics for 208 countries, per capita income ranking for the 21 countries was obtained. Per capita income in descending order ranged from 5th (United States) to 132nd (China).

The 21 school systems from which teacher reports were obtained differ in many ways. According to UNESCO statistics (UNESCO Institute for Statistics, 2006), the number of years of compulsory schooling ranges from 5 in Iran to more than 10 in Japan, Australia, Europe, and the United States. Education expenditure as a percentage of gross national product varies widely, from 3% in Lebanon to 9% in Denmark. There are also large differences in pupil–teacher ratios in primary schools, ranging from 10 or 11 in Greece, Italy, and Portugal to 30 in Jamaica. In addition, the countries vary in school organization, models of pedagogy, and curriculum.

The number of students rated by each teacher varied across countries. In 5 countries, teachers rated all the students in their class. In the other 16 countries, teachers were instructed to rate a prescribed number of students (3–8 students per class in 3 countries, and $\leq$2 students per class in 13 countries), with the researchers selecting participants using a randomization algorithm (e.g., the seventh student on an alphabetical list for each class, or three boys and three girls per class based on a random number table).

Response rates varied across samples, with a range from 72% to 100%. The response rate for each sample was computed by dividing the total number of completed TRFs by the total number given to teachers. Finally, composition of the sample varied across countries. Specifically, 7 of the 21 samples excluded students known to have been referred for mental health or special education services in the preceding year, whereas 14 did not.

All 21 of our samples met Nunnally and Bernstein's (1994) recommended minimum *N* of 300 per sample, even after TRFs were excluded because ratings were missing for >8 problem items, which is the standard criterion for excluding TRFs. For 17 countries, from 0% to 5% of TRFs were excluded for >8 missing items, whereas four countries had higher percentages of TRFs with >8 missing items (Iran 7%, Lebanon 13%, Hong Kong 15%, and Romania 22%). TRFs were also excluded because of missing age or gender information in Hong Kong (5%), Iran (0.5%), and Lebanon (0.2%).

For purposes of this study, we analyzed data only for children from 6 to 15 years of age. Because many countries have a school-leaving age of 16, school-based samples that include students $\geq$16 years might disproportionately exclude dropouts, who might have elevated problem scores. Our final sample

Table 1
Reference, *N*, Ages, Sampling Procedure, Per Capita Income Rank, Number of TRFs Completed Per Teacher, Response Rate, and Exclusion of Referred Children for 21 Countries

Country	Reference	*N*	Ages	Sampling Procedure	Per Capita Income Rank	TRFs Per Teacher	Response Rate	Referred Excluded?
Australia	Zubrick et al., 1997	1,697	6–15	Household-based Western Australia random sample stratified by region; maximum of 2 siblings per family used in this study	24	≤2	78%	No
China	Liu et al., 2000a, 2000b	4,857	6–15	Household-based sample in 6 districts in Shandong Province stratified by SES based on municipal household registers	132	≤2	92%	No
Denmark	Bilenberg, 1999	599	6–15	Regional stratified random general population sample from household registers in Fynen, one of the main Danish islands	6	≤2	84%	No
Finland	Weintraub et al., 2004	1,695	6–15	Regional school-based sample; recruitment of all children in 134 classes of 8 public schools in southern Finland	14	Whole class	92%	No
France	Fombonne & Vermeersch, 1997	493	6–11	Epidemiological sample of children of employees of national utility, one child per family; TRFs sent and returned by mail	19	≤2	Unknown	Yes
Greece	Roussos et al., 1999	1,179	6–12	National school-based random sample, 5 children per grade in 6 grades of 42 schools	42	≤8	99%	No
Hong Kong	Leung et al., 2006	1,993	6–15	City school-based sample in 89 schools	25	≤2	83%	No
Iran	Minaei, 2005	1,025	6–15	School-based sample in Tehran region; 10 students in each of 180 schools	111	≤2	100%	Yes
Italy	Frigerio et al., 2004	1,022	6–15	Regional random school-based sample in 31 municipalities of northern Italy; one class per grade, in 61 public schools	26	≤8	79%	Yes
Jamaica	Lambert & Lyubansky, 1999	615	6–15	Random school-based sample in 45 schools in two regions	98	≤2	87%	Yes

Table 1 (*Continued*)

Country	Reference	*N*	Ages	Sampling Procedure	Per Capita Income Rank	TRFs Per Teacher	Response Rate	Referred Excluded?
Japan	Kanbayashi, 2000	2,559	6–15	National sample in 72 schools in 10 prefectures	9	≤8	82%	No
Lebanon	Bathiche, 2005	1,504	6–11	School-based sample; 121 teachers in 18 schools	75	Whole class	95%	No
Lithuania	Zukauskiene & Kajokiene, 2004	2,601	6–15	National school-based sample; 2 classes per grade in 153 schools	73	≤2	84%	No
Netherlands	Verhulst et al., 1997	1,239	6–15	National probability sample of Dutch children drawn from household registers (89 areas); TRFs mailed	16	≤2	83%	No
Poland	Wolanczyk, 2003	2,133	6–15	National school-based sample in 132 schools in 17 regions, one class per grade sampled	72	Whole class	99%	No
Portugal	Fonseca, 1995	1,373	6–15	Regional stratified school-based sample in Coimbra; classes in Grades 2, 4, 6	48	Whole class	91%	No
Puerto Rico	Achenbach et al., 1990	565	4–16	Islandwide household sample in 210 population clusters; interviewers administered TRFs	50	≤2	93%	No
Romania	Domuta, 2004	922	6–11	School-based sample in four towns or regions	97	Whole class	80%	Yes
Thailand	Weisz et al., 1989	359	6–11	National school-based sample; 38 schools, randomly selected classes	104	1	92%	Yes
Turkey	Erol & Simsek, 1997	1,600	6–15	National general population household sample; interviewers administered TRFs	90	≤2	88%	Yes
United States	Achenbach & Rescorla, 2001	927	6–15	National general population sample of households in 40 states; TRFs mailed	5	≤2	72%	No

Note: TRF = Teacher's Report Form; SES = socioeconomic status.

sizes ranged from 359 for Thailand to 4,857 for China. Three age brackets were used for analysis: 6–8, 9–11, and 12–15 years. We used age brackets rather than analyzing age as a continuous variable so that we could compare TRF scores for children within specific developmental periods. We required a minimum of 80 students per country per age bracket. Because all 21 samples could be analyzed for ages 6–11 (N = 20,664), these analyses provide the primary results reported here. We also report supplementary results from analyses of ages 6–15 in 16 of the countries (N = 26,500), when they provide important additions to our findings.

Assessment Instrument

Except for Jamaica, Australia, and the United States, translations of the TRF were used. As described earlier, the typical translation procedure was to have professional linguists or translators translate the TRF. Back-translations were then done to ensure that the translation accurately captured the meaning of the English original. The translators and researchers inspected each translated problem item to ensure that it accurately captured the meaning of the English original, made sense, and would be understood by teachers.

In 17 of the 21 countries, either the 1986 (Achenbach & Edelbrock, 1986) or the 1991 (Achenbach, 1991) edition of the TRF was used. In Iran, Lithuania, Romania, and the United States, the 2001 edition was used (Achenbach & Rescorla, 2001). No items were replaced when the 1986 TRF was revised in 1991, but Problem Items 5, 28, and 99 were replaced by new items for the 2001 TRF. These 3 items, as well as the open-ended Items 56h and 113, were therefore omitted from all analyses, resulting in the same 115 items for analysis in each of the 21 countries. Scores on the 2001 versions of TRF scales (minus Item 5 for the Withdrawn/Depressed syndrome and Items 28 and 99 for the Rule-Breaking Behavior syndrome) were computed from the same 115 items for all 21 samples. Item ratings (0, 1, or 2) were summed to yield scores for the eight syndromes plus the two empirically based Inattention and Hyperactivity-Impulsivity subscales; six DSM-oriented scales plus the two DSM-oriented Inattention and Hyperactivity-Impulsivity subscales; and Internalizing, Externalizing, and Total Problems scales (21 scales in all).

Data Analysis

For the following analyses, we used the complete sample from every country (N = 30,957): (a) to test internal consistency reliability (alpha coefficients) for each TRF scale; (b) to determine the overall mean for Total Problems scores, which we calculated by averaging the mean Total Problems scores for the 21 countries, to avoid weighting the overall mean by differences in sample sizes; (c) to compute correlations between the mean item ratings from each country with the mean item ratings from every other country; and (d) to compute a multilevel regression testing the effects of per capita income ranking, number of students rated by each teacher, response rates, and exclusion of referred students on Total Problems scores.

To analyze country, gender, and age effects on Total Problems, we used a 21 (country) × 2 (gender) × 2 (ages 6–8, 9–11) ANOVA for the 21 samples containing ≥80 students in each of the two age brackets. Five 21 × 2 × 2 multivariate ANOVAs (MANOVAs) were used to analyze scores for Internalizing and Externalizing, for the eight syndromes, for the empirically based Inattention and Hyperactivity-Impulsivity subscales, for the six DSM-oriented scales, and for the DSM-oriented Inattention and Hyperactivity-Impulsivity subscales. Within-country gender and age effects were tested by separate Gender × Age ANOVAs for each scale that had a significant gender and/or age effect in the Country × Gender × Age analyses. Parallel analyses were conducted for ages 6–15 in 16 countries, except that three rather than two age brackets were used (6–8, 9–11, and 12–15).

TRF scores in every country were positively skewed (i.e., had more low than high scores), as is typical for problem scores in general population samples. This is because

most students have minimal problems and relatively few have numerous problems. Transformations do not make distributions of such scores normal. For example, when logarithmic and square-root transformations were applied to Total Problems scores, both skewness and kurtosis were still significant. However, the ANOVA and MANOVA models used in this study were very robust with respect to deviations from normality, especially with large samples (Kirk, 1995). Accordingly, as in previous studies using the TRF, untransformed raw scores were used in all analyses. As would be expected because of the power afforded by such large samples, significant results for homogeneity tests were obtained in this study. However, the ANOVA and MANOVA models used were also robust with respect to heterogeneity of variances and covariances.

Because of the large number of analyses and the high statistical power, we used $p <$ .001 as the alpha level for ANOVAs, MANOVAs, and the bicountry correlations. For significant findings, ESs indicating the percent of variance accounted for in the ANOVAs and MANOVAs are reported in terms of η^2. F values, which are less informative, are not reported but can be obtained from the authors.

Results

Internal Consistency of TRF Scales Across Countries

Alpha coefficients for the 21 TRF scales analyzed for this study are summarized in Table 2. Across the 21 countries, the mean alpha for the 21 TRF scales was .82. Mean alphas for Total Problems, Internalizing, and Externalizing were .96, .86, and .92, respectively. Mean alphas ranged from .65 to .92 for the empirically based syndromes and subscales and from .64 to .91 for the DSM-oriented scales and subscales. TRF Total Problems, Externalizing, Attention Problems, Aggressive Behavior, and DSM-oriented Attention Deficit Hyperactivity Problems showed consistently high alphas across countries ($\geq$.91). When the 21 alphas obtained for each country were correlated with the 21 alphas for each of the other countries, all 210 bicountry correlations were significant at $p <$ .001. The mean of the 210 bicountry correlations was .90, with a range from .65 (Greece with Thailand) to .99 (Japan with Australia). Only 11 of the 210 bicountry correlations were <.80, 9 of which involved Thailand, which was also the smallest sample. Thus, the scales manifested very similar internal consistencies across the countries.

Total Problems Scores

Although Total Problems scores could range from 0 to 230 (i.e., if all 115 items were rated 2), mean Total Problems scores ranged only from 8.4 for Japan to 31.6 for Puerto Rico, as illustrated in Figure 1. Fifteen countries differing widely in language, ethnicity, geographic region, religion, political and economic system, and size scored within one *SD* ($\pm$6.2) of the overall mean of 21.6 (15.1–27.2). For example, the United States, Lebanon, Hong Kong, and Portugal had Total Problems scores in the narrow range of 19.6–22.0. Japan, Finland, and China scored more than 1 *SD* below the overall mean, whereas Jamaica, Puerto Rico, and Thailand scored more than 1 *SD* above this mean.

Table 3 displays significant ESs for country, gender, age, and interactions for the 21-country analyses for ages 6–11. Findings for the 16-country ages 6–15 analyses appear in parentheses. Table 4 displays the number of significant within-country gender differences for each scale that had a significant gender effect in the Country $\times$ Gender $\times$ Age analyses. In presenting results, we focus on ages 6–11, only describing findings for ages 6–15 when they differ in notable ways.

As shown in Table 3, the ages 6–11 ANOVA of Total Problems had a country ES of 8% and a gender ES of 2%. There was also a significant Gender $\times$ Country interaction effect on Total Problems (ES < 1%), reflecting slightly larger gender differences in some countries than in others. As shown in Table 4, boys obtained significantly higher TRF Total Problems scores than girls in 18 of the 21 countries, but the within-country ANOVAs

Table 2
Internal Consistency Alpha Coefficients for TRF Scales Across 21 Countries

TRF Scale	Minimum	Maximum	Mean
Total Problems	.94	.97	.96
Internalizing	.82	.89	.86
Externalizing	.86	.95	.92
Syndromes			
Anxious/Depressed	.72	.85	.80
Withdrawn/Depressed	.69	.87	.78
Somatic Complaints	.50	.83	.68
Social Problems	.61	.84	.75
Thought Problems	.53	.78	.65
Attention Problems	.89	.95	.93
Inattention Subscale	.85	.95	.92
Hyperactivity/Impulsivity Subscale	.80	.93	.90
Rule-Breaking Behavior	.50	.83	.72
Aggressive Behavior	.83	.94	.92
DSM-Oriented Scales			
Affective Problems	.61	.78	.70
Anxiety Problems	.48	.72	.64
Somatic Problems	.51	.79	.66
Attention Deficit Hyperactivity	.85	.94	.91
Inattention Subscale	.74	.91	.87
Hyperactive Impulsive Subscale	.77	.93	.90
Oppositional Defiant Problems	.51	.88	.80
Conduct Problems	.72	.89	.86
Mean alpha across all scales	**.72**	**.85**	**.82**

Note: TRF = Teacher's Report Form; DSM = *Diagnostic and Statistical Manual.*

were not significant in the United States, Jamaica, or Iran. There was no significant effect of age on Total Problems, but the Age × Country interaction was significant (ES < 1%). Within-country ANOVAs yielded significant age effects on Total Problems in 4 of 21 countries. Students ages 6–8 scored higher in China and Finland and lower in Jamaica and Romania than students ages 9–11. When Total Problems scores were analyzed for ages 6–15 in 16 countries, the results were very similar to those for ages 6 –11 in 21 countries.

Internalizing and Externalizing Scores

The 21-country MANOVA for ages 6–11 yielded an ES of 11% for country on Internalizing (see Table 3). Japan had the lowest scores, whereas Puerto Rico, Iran, and Thailand had the highest scores. The gender effect was not significant. Age, Gender × Country, and Age × Country effects were significant but <1%. Within-country ANOVAs yielded no significant effects for age, gender, or Age × Gender.

The 21-country ages 6–11 MANOVA yielded an ES of 4% for country on Externalizing (see Table 3). Japan had the lowest mean score, whereas Iran and Jamaica had the highest mean scores. The ES for gender was 3%. Boys scored significantly higher than girls on Externalizing in 17 of 21 countries (not in Jamaica, Puerto Rico, the United States, or Iran; see Table 4). The overall age effect was not significant.

When Internalizing and Externalizing scores were analyzed for ages 6–15 in 16

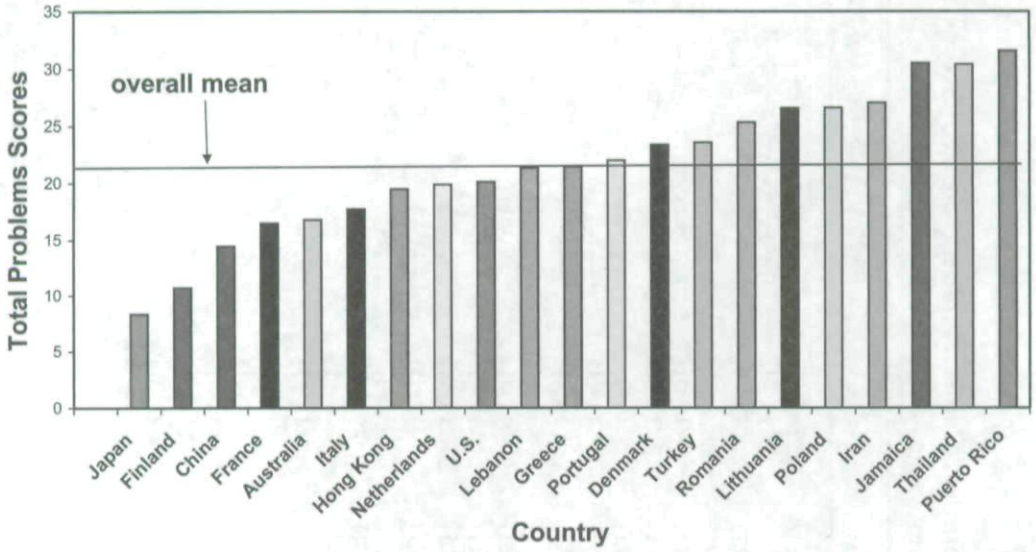

Figure 1. TRF Total Problems scores in 21 countries for ages 6–15 (N = 30,957).

countries, the results were very similar to those for ages 6–11 in 21 countries. A noteworthy supplemental result was that the overall gender effect on Internalizing was significant for ages 6–15. However, girls obtained significantly higher scores than boys in only 2 of 16 countries.

Syndrome Scores

In the 21-country ages 6–11 MANOVA of the eight syndromes, ESs for country ranged from 4% (Somatic Complaints, Thought Problems, and Rule-Breaking Behavior) to 13% (Anxious/Depressed; see Table 3). Gender effects were significant for five syndromes (Social Problems, Thought Problems, Attention Problems, Rule-Breaking Behavior, and Aggressive Behavior), with ESs ranging from <1% to 4%. The 21 within-country ANOVAs indicated that boys scored significantly higher than girls in 7 countries for Social Problems, 12 countries for Thought Problems, 19 countries for Attention Problems, 17 countries for Rule-Breaking Behavior, and 16 countries for Aggressive Behavior (see Table 4). Figure 2 vividly illustrates the gender disparity for Attention Problems in 19 of the 21 countries (not Jamaica or Iran). Age effects were significant for only two syndromes (Somatic Complaints and Rule-Breaking Behavior).

The 16-country ages 6–15 MANOVA yielded results similar to those for ages 6–11. A notable difference in these supplementary analyses was that there were significant gender effects for Anxious/Depressed and Somatic Complaints (both ESs <1%), with girls scoring higher than boys in 3 of 16 and 2 of 16 countries, respectively. Significant age effects were more numerous for ages 6–15 than for ages 6–11, but all ES were <1%. Rule-Breaking Behavior had the most consistent age effect (in 7 of 16 countries), with highest scores generally being for ages 12–15.

DSM-Oriented Scales

In the 21-country MANOVA for DSM-oriented scales using children from 6–11 years of age, country ESs ranged from 3% for Somatic Problems to 8% for Anxiety Problems (see Table 3). Gender differences ranged from <1% to 5%. Girls scored significantly higher than boys in 3 of 21 countries on Somatic Problems. For Affective Problems, the within-country gender difference was significant only in Finland, Italy, Japan, and Thailand. Surpris-

Table 3

Effect Sizes (η^2) Associated With Significant Differences as a Function of Country, Gender, and Age on TRF Scores for 20,664 Students Ages 6–11 from 21 Countries and for 26,500 Students Ages 6–15 From 16 Countries

Scale	Country	Gender	Age	C × G	C × A	G × A	C × G × A
Total Problems	8 (9)[a]	2 (2)	—[b] (—)	<1 (<1)	<1 (<1)	— (—)	— (—)
Internalizing	11 (12)	— (<1)	<1 (<1)	— (<1)	<1 (<1)	— (—)	— (—)
Externalizing	4 (5)	3 (2)	— (—)	<1 (<1)	<1 (<1)	— (—)	— (—)
Syndromes							
Anxious/Depressed	13 (13)	— (<1)	— (—)	<1 (<1)	<1 (<1)	— (<1)	— (—)
Withdrawn/Depressed	6 (8)	— (—)	— (<1)	<1 (<1)	<1 (<1)	— (—)	— (—)
Somatic Complaints	4 (4)	— (<1)	<1 (—)	<1 (<1)	<1 (1)	— (—)	— (—)
Social Problems	5 (5)	<1 (<1)	— (<1)	<1 (<1)	<1 (<1)	— (—)	— (—)
Thought Problems	4 (3)	<1 (<1)	— (<1)	— (<1)	<1 (<1)	— (<1)	— (<1)
Attention Problems	5 (6)	4 (4)	— (—)	<1 (<1)	<1 (<1)	<1 (<1)	— (—)
Inattention Subscale	5 (6)	2 (3)	— (<1)	<1 (<1)	<1 (<1)	<1 (<1)	— (—)
Hyperactivity/Impulsivity Subscale	5 (5)	5 (5)	— (<1)	<1 (<1)	<1 (<1)	— (—)	— (—)
Rule-Breaking Behavior	4 (4)	2 (2)	<1 (<1)	<1 (<1)	<1 (<1)	— (—)	— (—)
Aggressive Behavior	4 (4)	3 (2)	— (—)	<1 (<1)	<1 (<1)	— (—)	— (—)
DSM-Oriented Scales							
Affective Problems	4 (5)	<1 (<1)	<1 (<1)	— (<1)	— (<1)	— (—)	— (—)
Anxiety Problems	8 (8)	— (—)	— (—)	— (—)	— (<1)	— (—)	— (—)
Somatic Problems	3 (3)	<1 (<1)	<1 (—)	<1 (<1)	<1 (<1)	— (—)	— (—)
Attention Deficit Hyperactivity Problems	5 (5)	5 (4)	— (<1)	<1 (<1)	<1 (<1)	<1 (<1)	— (—)
Inattention Subscale	4 (5)	3 (3)	— (—)	<1 (<1)	— (<1)	<1 (—)	— (—)
Hyperactivity/Impulsivity Subscale	4 (5)	5 (4)	<1 (<1)	<1 (<1)	<1 (<1)	— (—)	— (—)
Oppositional Defiant	3 (3)	2 (2)	— (<1)	<1 (<1)	<1 (<1)	— (—)	— (—)
Conduct Problems	3 (3)	4 (3)	<1 (—)	<1 (<1)	<1 (<1)	— (—)	— (—)

Note: TRF = Teacher's Report Form; C = country; G = gender; A = age.

[a] Parentheses indicate effect sizes for ages 6–15 in 16 countries.

[b] Dashes indicate effect was not significant at $p \leq .001$.

Table 4
Number of Countries With Significant Within-Country Gender Effects by TRF Scale

Scale	Ages 6–11, 21 Countries	Ages 6–15, 16 Countries
Total Problems	18 B > G	14 B > G
Internalizing	[a]	2 G > B
Externalizing	17 B > G	14 B > G
Syndromes		
Anxious/Depressed	[a]	3 G > B
Withdrawn/Depressed	[a]	[a]
Somatic Complaints	[a]	2 G > B
Social Problems	7 B > G	6 B > G
Thought Problems	12 B > G	11 B > G
Attention Problems	19 B > G	15 B > G
Inattention subscale	19 B > G	14 B > G
Hyperactivity/Impulsivity subscale	19 B > G	15 B > G
Rule-Breaking Behavior	17 B > G	13 B > G
Aggressive Behavior	16 B > G	14 B > G
DSM-Oriented Scales		
Affective Problems	4 B > G	4 B > G
Anxiety Problems	[a]	[a]
Somatic Problems	3 G > B	3 G > B
Attention Deficit Hyperactivity Problems	19 B > G	15 B > G
Inattention subscale	19 B > G	14 B > G
Hyperactivity/Impulsivity subscale	19 B > G	15 B > G
Oppositional Defiant Problems	17 B > G	14 B > G
Conduct Problems	18 B > G	15 B > G

Note: TRF = Teacher's Report Form; B = boy; G = girl.
[a] Gender effect not significant in multicultural analyses.

ingly, boys obtained higher scores than girls in all four countries. With much more consistency across countries, boys scored significantly higher than girls on Attention Deficit Hyperactivity Problems (19 of 21 countries, not Jamaica or Iran), Oppositional Defiant Problems (17 of 21 countries, not Jamaica, Puerto Rico, Iran, or the United States), and Conduct Problems (18 of 21 countries, not Jamaica, Puerto Rico, or Iran). Age effects were significant for the DSM-oriented Affective Problems, Somatic Problems, and Conduct Problems scales, with older students scoring higher than younger students (ESs < 1%). However, within-country ANOVAS indicated very few significant age effects on the DSM-oriented scales. Results of the 16-country ages 6–15 MANOVA of DSM-oriented scale scores were very similar to those obtained for ages 6–11.

Inattention and Hyperactivity-Impulsivity Subscales

In the 21-country, ages 6–11 MANOVA of the empirically based Attention Problems subscales, ESs for country were 5% for both the Inattention and the Hyperactivity-Impulsivity subscales (see Table 3). The gender ES was larger for the Hyperactivity-Impulsivity subscale (5%) than for the Inattention subscale (2%). The age effect was not significant for either subscale. For both subscales, boys scored significantly higher than girls in 19 of 21 countries (not Jamaica or Iran; see Table 4). Results for the DSM-oriented subscales

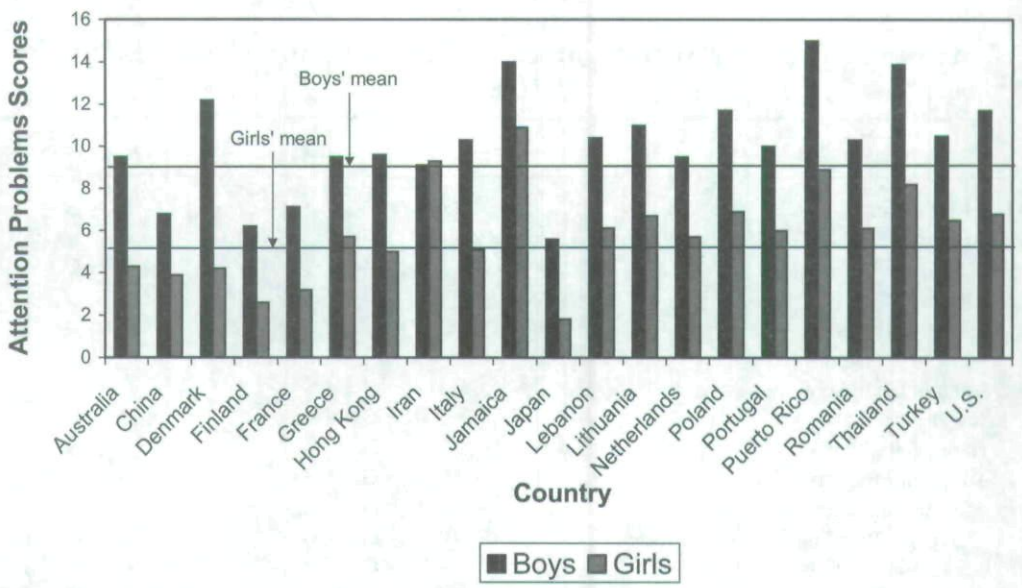

Figure 2. TRF Attention Problems scores by gender in 21 countries for ages 6–11 (N = 20,664).

were very similar, as were the results for ages 6–15 in 16 cultures.

Mean Item Ratings

After the mean 0–1–2 rating for each of the 115 problem items within each country was computed, correlations were computed between the 115 mean item ratings obtained in each country and the 115 mean item ratings in each of the other 20 countries. This yielded a total of 210 bicountry correlations, all of which were significant at $p < .001$. The mean of these 210 correlations was .74. Bicountry correlations ranged from .53 (China with France) to .97 (United States with Australia). Each country's 20 bicountry correlations were averaged to yield a mean *r*-value per country. The following 11 of the 21 countries had mean values of $r \geq .75$ across all pairwise comparisons: Australia, Denmark, Greece, Hong Kong, Italy, Jamaica, Lebanon, the Netherlands, Poland, Portugal, and the United States. China had the lowest mean *r* value (.65), whereas Lebanon had the highest (.80).

Five of the 10 TRF items with the highest mean ratings across the 21 countries were from the Attention Problems syndrome: (item 4) Fails to finish things he/she starts; (8) Can't concentrate, can't pay attention for long; (61) Poor schoolwork; (78) Inattentive or easily distracted; and (92) Underachieving, not working up to potential. The remaining 5 were all Internalizing items: (item 32) Feels he/she has to be perfect; (71) Self-conscious or easily embarrassed; (75) Too shy or timid; (81) Feels hurt when criticized; and (108) Is afraid of making mistakes. The 10 items with the lowest mean ratings were as follows: (item 18) Deliberately harms self or attempts suicide; (40) Hears sounds or voices that aren't there; (56a) Aches or pains; (56e) Rashes or other skin problems; (56g) Vomiting, throwing up; (70) Sees things that aren't there; (82) Steals; (85) Strange ideas; (91) Talks about killing self; and (105) Uses alcohol or drugs for nonmedical purposes.

Per Capita Income, Respondent Burden, Response Rates, and Exclusion of Referred Students

For the complete sample ($N = 30,957$), we used multilevel random effects regression implemented by Mplus (Muthén & Muthén, 2001) to test the effects of the following vari-

ables on Total Problems scores: per capita income ranking, number of students rated by each teacher, response rate, and exclusion of referred students. None of the predictors was significant (all $p > .10$). However, the correlation between per capita income ranking and Total Problems ($r = .53$, $p < .05$) indicated a tendency for the poorest countries to have the highest problem scores when the effects of other variables were not partialed out. Two of the three countries with the lowest Total Problems scores ranked high in per capita income (Japan, ranked 9th; Finland, ranked 14th), but the third (China) had the lowest per capita income of the 21 countries (ranked 132nd). Two of the three countries with the highest Total Problems scores had per capita incomes that were quite low (Jamaica, ranked 98th; Thailand, ranked 104th), but the third (Puerto Rico) did not (ranked 50th).

Discussion

To our knowledge, this multinational study is the first to report systematic comparisons of teacher-reported problems on the same instrument across more than a few countries. Although the 21 countries in our study (10 in Europe, 4 in Asia, 3 in the Middle East, 2 in the Caribbean, plus the United States and Australia) vary widely in language, ethnicity, geographic region, religion, political and economic system, educational system, and size, many consistencies were found across countries.

Despite the vicissitudes of translation, TRF scales showed similar internal consistency across the 21 countries. The mean bicountry r value of .90 for alphas indicates strong cross-national replication for this important aspect of construct validity. The high bicountry correlations between alphas complement confirmatory factor results indicating good fit for the factor model for these same data reported by Ivanova et al. (2006). Across the countries, the highest alphas were obtained for Total Problems, Internalizing, Externalizing, Attention Problems, Aggressive Behavior, and DSM-oriented Attention Deficit Hyperactivity Problems.

Fifteen of the 21 samples had mean Total Problems scores within 1.0 *SD* (6.2) of the overall mean of 21.6. These 15 samples included 9 in Europe, 3 in the Middle East, plus Hong Kong, the United States, and Australia. Two Asian samples (Japan and China) and one Northern European sample (Finland) scored >1 *SD* below the overall mean on Total Problems, whereas two Caribbean samples (Puerto Rico and Jamaica) and Thailand scored >1 *SD* above the overall mean. No obvious common feature is shared by the countries with either the low-scoring or the high-scoring samples.

Of the 42 significant ESs for country, 28 were small (1% to 5.9%), 14 were medium (5.9–13.8%), and none were large by Cohen's (1988) benchmarks. The largest ES for country was for the Anxious/Depressed syndrome (13%). Greece, Turkey, Thailand, and Iran had the highest Anxious/Depressed scores, whereas Japan, Finland, and China had the lowest.

Previous TRF studies have demonstrated significant within-country associations between SES and TRF problem scores (Achenbach & Rescorla, 2001; Achenbach et al., 1987, 1990; Bilenberg, 1999; Fonseca et al., 1995; Lambert et al., 1995; Roussos et al., 1999). Although per capita income was significantly correlated with Total Problems scores in our study, it was not a significant predictor when other variables were partialed out in the multilevel regression analysis.

Response rate was not significantly correlated with Total Problems scores ($r = .29$), nor was it a significant predictor in the multilevel regression. Some countries with high problem scores had very high response rates (e.g., Puerto Rico) and some countries with low problem scores had low response rates (e.g., Finland). Yet, there were many exceptions to this general trend, most notably China and Japan, where response rates were high but problem scores were low.

It might be expected that scores would be lower when teachers rated all the students in their class than when they rated just a few randomly chosen students. However, the number of students rated by each teacher was not significantly associated with Total Problems

scores. In addition, inclusion of referred students was not significantly associated with Total Problems scores, although the number of referred students was generally small.

Comparisons of TRF Findings with CBCL Findings

Rescorla et al. (2006a) compared CBCL scores from 31 countries (N = 55,508 children), 20 of which provided TRF data for the present study. The overall mean Total Problems scores for the TRF and the CBCL were very similar (21.6 vs. 22.5). On both the TRF and the CBCL, scores in Japan and China tended to be low, whereas scores in Puerto Rico tended to be high. On the other hand, Finland, Jamaica, and Thailand had scores within 1.0 *SD* of the overall mean on the CBCL, whereas Finland had low scores and Jamaica and Thailand had high scores on the TRF.

CBCL ESs for country were 3–14%, consistent with TRF ESs. For both the TRF and the CBCL, Anxious/Depressed was the syndrome with the largest ES for country. This suggests that parents and teachers in different countries vary more in how they perceive problems associated with negative affectivity than in how they perceive aggressive, disruptive, and oppositional behaviors.

TRF gender effects were very consistent across the 21 countries for Externalizing and Attention Problems, with boys scoring significantly higher than girls in most countries. Similar findings were obtained for the CBCL (Rescorla et al., 2006a). However, gender differences on the TRF were less consistent across the 21 countries for Internalizing and its three constituent syndromes, with no significant within-country gender effects for ages 6–11. Even when the sample included ages 12–15, girls obtained significantly higher TRF Internalizing scores than boys in only a few countries. By contrast, girls ages 6–16 obtained significantly higher scores than boys on the CBCL Internalizing, Anxious/Depressed, and Somatic Complaints scales in 8, 9, and 11 out of 24 countries, respectively (Rescorla et al., 2006a).

This pattern of results indicates that, across many countries, both teachers and parents rate boys as having more attention and externalizing problems than girls. On the other hand, parents are more likely than teachers to rate girls as having more internalizing problems than boys. This may be because externalizing problems are more easily observed than internalizing problems by both parents and teachers, and as suggested by other research, boys generally engage in more externalizing behaviors than girls. Teachers, however, may not be aware of internalizing problems experienced by either boys or girls, because students may not divulge their feelings to teachers. This would mask any underlying gender difference on internalizing in teacher-reported problems. Parents, on the other hand, may be more privy to girls' internalizing problems because their daughters may express them more openly at home. This interpretation is supported by findings of even greater gender differences on the Internalizing scale on the YSR than on the CBCL (Rescorla, Almqvist et al., 2006b).

The very few within-country age effects on TRF scores varied in direction. The most consistent age trends were increases in Rule-Breaking Behavior scores with age in 7 countries. Age effects were somewhat more consistent on the CBCL (Rescorla et al., 2006a), although all ESs were $\leq$1%. Scores on the CBCL Internalizing scale tended to increase with age, whereas scores on the CBCL Externalizing scale tended to decrease with age.

Limitations

Although TRFs were obtained for broad school samples in each of 21 countries, some methodological differences between the countries may constitute possible limitations. The countries differed in whether referred students were excluded, in teachers' response rates, and in the number of TRFs completed by each teacher. However, these factors were not significantly associated with Total Problems scores. Another limitation is that SES information was not provided for students in enough countries to be analyzed.

Although the multistep process of translation, back-translation, and review of items was typically followed in the countries providing data for this study, details of the translation process varied. In addition, translations into languages closely related to English, such as German, Dutch, French, Spanish, or Italian, may have yielded closer equivalents to U.S. items than translations into languages such as Polish, Lithuanian, Russian, Arabic, Farsi, Chinese, Turkish, or Japanese. Possible variations in the quality of TRF translations constitute another limitation of the study.

The methodology of this study did not test how teachers in the different countries interpreted the TRF items or whether, in fact, they even perceived these items as representing "problems." For example, the fact that Item 78 (Inattentive or easily distracted) received high scores in almost every culture does not prove that teachers in every culture viewed being inattentive as problematic. To explore this issue, research using an *emic* approach is needed, whereby the meanings of the constructs studied would be explored within each cultural context. Thus, a possible limitation of this study is its *etic* approach, whereby the same standardized assessment instrument was used in 21 countries.

In addition, because our items do not tap all possible problems teachers might report, use of additional problem items might yield different findings. However, our study does provide systematic comparisons of scores on the same items and scales for students in 21 countries that differ widely in ethnicity, geographical region, religion, economic and political systems, and historical experience.

A final limitation of the study is that the lowest alphas found for five of the syndromes and four of the DSM-oriented scales were $<.70$, indicating modest internal consistency in some countries. However, mean alphas for all syndromes and DSM-oriented scales were $\geq.72$, indicating adequate internal consistency overall. Furthermore, the mean bicountry correlation of .90 for alphas indicated great consistency across countries with respect to which TRF scales had excellent, moderate, or modest alphas.

Implications of the Findings

Despite many differences among the 21 countries in language, ethnicity, geographic region, population, size, religion, political and economic system, per capita income, class size, educational funding, curriculum, and educational philosophy, the TRF yielded quite similar results. The significant ESs for country clearly indicated some effect on TRF scores. However, country rarely explained more than 10% of the variance in TRF scores. Furthermore, differences between countries in teacher-reported problems were mostly of small magnitude. Internal consistency reliability was very similar across countries, as were gender effects and mean item ratings. On balance, TRF scores varied much more within countries than between countries.

Practical Implications

The 10 TRF items with the highest mean scores across all 21 countries highlight problems that are salient to teachers despite differences in languages, cultures, and educational systems. These included poor attention, concentration, task persistence, and school achievement, as well as shyness, perfectionism, being self-conscious, being afraid to make mistakes, and being overly sensitive to criticism. The 10 items with the lowest scores across all 21 countries reflect problems that are less commonly observed in the school setting, such as seeing and hearing things that are not there, strange ideas, suicide or self-injury, stealing, substance use, and various somatic problems without known medical causes.

A reasonable inference from the data presented is that teachers in most countries expect students to pay attention, complete their work, maintain self-control, accept criticism, deal with new situations, and get along with peers. Clearly, difficulties in these areas were among the most commonly reported by teachers in 21 countries. Furthermore, problems suggestive of more significant psychopathology (e.g., strange thoughts and actions, self-injury, stealing, substance use, and delinquent activity), were rarely reported by teach-

ers. Thus, such problems may justify major concern when teachers do report them.

The high internal consistency of TRF scales in many countries and the similarities of mean scale scores, mean item ratings, and age and gender findings attest to the multicultural robustness of the TRF. The more research findings demonstrate that an instrument's results can be generalized to different cultural contexts, the more confident school psychologists can be in using that instrument and in interpreting its findings for diverse students.

References

Achenbach, T. M. (1991). *Manual for the Teacher's Report Form and 1991 Profile.* Burlington, VT: University of Vermont, Department of Psychiatry.

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

Achenbach, T. M., Dumenci, L., & Rescorla, L. A. (2003). DSM-oriented and empirically based approaches to constructing scales from the same item pools. *Journal of Clinical Child and Adolescent Psychology, 32,* 328–340.

Achenbach, T. M., & Edelbrock, C. (1986). *Manual for the Teacher's Report Form and Teacher Version of the Child Behavior Profile.* Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms & Profiles.* Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families.

Achenbach, T. M., & Rescorla, L. A. (2007) *Multicultural understanding of child and adolescent psychopathology: Implications for mental health assessment.* New York: Guilford Press.

Achenbach, T. M., Verhulst, F. C., Edelbrock, C., Baron, G. D., & Akkerhuis, M. S. (1987). Epidemiological comparisons of American and Dutch children: II. Behavioral/emotional problems reported by teachers for ages 6 to 11. *Journal of the American Academy of Child and Adolescent Psychiatry, 27,* 326–332.

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

Bathiche, M. (2005). *The prevalence and validity of ADHD symptoms in Lebanon.* Unpublished doctoral dissertation, McGill University, Montreal, Canada.

Bilenberg, N. (1999). The Child Behavior Checklist (CBCL) and related material: Standardization and validation in Danish population and clinically based samples. *Acta Psychiatrica Scandivica, Supplementum, 100, 398,* 1–52.

Bracken, B., & Barona, A. (1991). State of the art procedures for translating, validating, and using psychoeducational tests in cross-cultural assessment. *School Psychology International, 12,* 119–132.

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

Cronbach, L. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

Domuta, A. (2004). *Evaluarea sindromului ADHD la varsta prescolara.* Teza de doctorat. Universitatea Baves-Bolyai, Cluj Napoca, Romania.

Dumenci, L., McConaughy, S. H., & Achenbach, T. M. (2004). A hierarchical three-factor model of inattention-hyperactivity-impulsivity derived from the attention problems syndrome of the Teacher's Report Form. *School Psychology Review, 33,* 287–301.

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

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

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

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

Geisinger, K. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment, 6,* 304–312.

Ivanova, M. Y., Achenbach, T. M., Dumenci, L., Rescorla, L. A., Almqvist, F., Bilenburg, N., et al. (2006). *Testing the Teacher's Report Form syndromes in 20 societies.* Manuscript submitted for publication.

Kanbayashi, Y. (2000). [Standardization data for the Japanese version of the Teacher's Report Form]. Unpublished raw data.

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

Lambert, M. C., Knight, F., Taylor, R., & Achenbach, T. M. (1995). Comparisons of behavioral and emotional problems among children of Jamaica and the United States: Teacher reports for ages 6–11. *Journal of Cross-Cultural Psychology, 27,* 82–97.

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

Leung, P. W. L., Kwong, S. L., Tang, C. P., Ho, T. P., Hung, S. F., Lee, C. C., et al. (2006). Test-retest reliability and criterion validity of the Chinese version of CBCL, TRF, and YSR. *Journal of Child Psychology and Psychiatry, 47,* 970–973.

Liu, X., Guo, C., Okawa, M., Zhai, J., Li, Y., Uchiyama, M., Neiderhiser, J., et al. (2000a). Behavioral and

emotional problems among Chinese children of divorce. *Journal of the American Academy of Child and Adolescent Psychiatry, 39,* 896–903.

Liu, X., Kurita, H., Guo, C., Tachimori, H., Ze, J., & Okawa, M. (2000b). Behavioral and emotional problems among Chinese children: Teacher reports for ages 6 through 11 years. *Journal of Child Psychology and Psychiatry, 41,* 253–260.

Minaei, A. (2005). *Manual of ASEBA School-Age Forms for Iranian Children.* Tehran: Research Institute for Exceptional Children.

Muthén, L. K., & Muthén, B. O. (1998–2004). *Mplus: User's guide.* Los Angeles: Muthén & Muthén.

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

Rescorla, L. A., Achenbach, T. M., Dumenci, L., Ivanova, M. Y., Almqvist, F., Bilenberg, N. et al. (2006a). Problems reported by parents in 31 countries. Manuscript submitted for publication.

Rescorla, L. A., Achenbach, T. M., Dumenci, L., Ivanova, M. Y., Almqvist, F., Bilenberg, N., et al. (2006b). Problems reported by adolescents in 24 countries. Manuscript submitted for publication.

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

UNESCO Institute for Statistics. (2006). *Education statistics.* Retrieved February 16, 2006, from http://www.uis.unesco.org

Verhulst, F. C., van der Ende, J., Ferdinand, R. F., & Kasius, M. C. (1997). The prevalence of DSM-III-R diagnoses in a national sample of Dutch adolescents. *Archives of General Psychiatry, 54,* 329–336.

Weine, A. M., Phillips, J. S., & Achenbach, T. M. (1995). Behavioral and emotional problems among Chinese and American children: Parent and teacher reports for ages 6–13. *Journal of Abnormal Child Psychology, 23,* 619–639.

Weintraub, S. (2004). [Standardization data for the Finnish version of the Teacher's Report Form]. Unpublished raw data.

Weisz, J. R., Suwanlert, S., Chaiyasit, W., Weiss, B., Weiss, Achenbach, T. M., et al. (1989). Epidemiology of behavioral and emotional problems among Thai and American children: Teacher reports for ages 6–11. *Journal of the American Academy of Child and Adolescent Psychiatry, 30,* 471–484.

Wolanczyk, T. (2003). Zaburzenia emocjonalne i behawioralne u dzieci i modziezy w Polsce. Warszawa: Akademia Medyczna w Warszawie.

World Bank. (2005). *World development indicators database, 15 July 2005.* Retrieved February 16, 2006 from http:/www.worldbank.org

Zukauskiene, R., & Kajokiene, I. (2004). *Child Behavior Checklist profile types in a random sample and in a child psychiatric clinic population in Lithuania.* Paper presented at Workshop on Child and Adolescent Assessment and Research Methodology, Vilnius, Lithuania, October 2004.

Zubrick, S. R., Silburn, S. R., Gurrin, L., Teoh, H., Shepherd, C., Carlton, J., et al. (1997). *Western Australian Child Health Survey: Education, health and competence.* Perth, WA: Australian Bureau of Statistics and the TVW Telethon Institute for Child Health Research.

Date Received: November 30, 2005
Date Accepted: June 30, 2006
Action Editor: Shane Jimerson ■

Leslie A. Rescorla received her PhD from Yale University. A licensed clinical psychologist and a certified school psychologist, she received her clinical training at the Yale Child Study Center, the Philadelphia Child Guidance Center, and the Children's Hospital of Philadelphia. She is Professor of Psychology and Director of the Child Study Institute at Bryn Mawr College. Her research interests focus on language delay in young children; empirically based assessment of emotional and behavioral problems in children, adolescents, and adults; and longitudinal patterns of school achievement.

Thomas. M. Achenbach is Professor of Psychiatry and Psychology and Director of the Center for Children, Youth and Families at the University of Vermont. He received his PhD at the University of Minnesota and was a postdoctoral fellow and faculty member at Yale University. His interests include developmental psychopathology, assessment of problems and adaptive functioning across the life span, and multicultural understanding of psychopathology.

Sofia Ginzburg is a doctoral student in the Clinical Developmental Psychology Program at Bryn Mawr College. She obtained an MEd from the University of Pennsylvania. She has worked as a school psychology intern in the Upper Darby School District, as a therapist in the Philadelphia public and parochial schools, and as a research coordinator at the University of Pennsylvania's Center for Health, Achievement, Neighborhoods, Growth and Ethnic Studies. Her dissertation research focuses on parent–adolescent agreement on adolescent behavioral and emotional problems in samples obtained from 20 countries.

Masha Y. Ivanova received her PhD from the State University of New York, Albany, and did a predoctoral clinical internship at the University of Vermont Center for Children, Youth, and Families. She is a Research Assistant Professor of Psychiatry at the University of Vermont. Her research focuses on the development of culturally sensitive assessments of children's emotional and behavioral functioning and family interventions to facilitate immigrant children's adjustment in the host culture.

Levent Dumenci received his PhD in psychometrics and statistics from Iowa State University in 1993. He is Research Associate Professor of Psychiatry at the University of Vermont. His primary interests include measurement models with categorical or continuous latent traits, structural equation modeling, measurement of change, and child psychopathology.

For brevity, biographical information is provided only for the first five authors.