School Psychology Review,
2007, Volume 36, No. 3, pp. 468–483

Testing the Teacher's Report Form Syndromes in 20 Societies

Masha Y. Ivanova,[1] Thomas M. Achenbach,[1] Leslie A. Rescorla,[2] Levent Dumenci,[1] 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] Patrick Leung,[13] Xianchen Liu,[14] Asghar Minaei,[15] Alexandra Roussos,[16] Zeynep Simsek,[17] Sheila Weintraub,[3] Tomasz Wolanczyk,[18] Stephen Zubrick,[19] Rita Zukauskiene,[20] and Frank C. Verhulst[21]

[1]University of Vermont, [2]Bryn Mawr College, [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]Chinese University of Hong Kong, [14]University of Pittsburgh, [15]Research Institute of Exceptional Children, Tehran, [16]Attiki Child Psychiatric Hospital, [17]University of Harran, [18]Medical University of Warsaw, [19]Curtin University of Technology and the Telethon Institute for Child Health Research, [20]Mykolas Romeris University, [21]Erasmus University Medical Center Sophia Children's Hospital—Sophia Children's Hospital

Abstract. Standardized assessment instruments developed in one society are often used in other societies. However, it is important to determine empirically how assessment instruments developed in one society function in others. The present study tested the fit of the Teacher's Report Form syndrome structures in 20 diverse societies using data for 30,030 6- to 15-year-old students from Asia; Australia; the Caribbean; eastern, western, southern, and northern Europe; and the Middle East. A correlated seven-syndrome model and a hierarchical Attention Problems model were tested separately in each of the 20 societies via confirmatory factor analyses. The results supported the fit of the models in the tested societies.

Standardized assessment instruments developed for children in one society are often used in other societies where the development of indigenous instruments may not be practical. However, before assessment instruments developed in one society can be applied in

Correspondence regarding this article should be addressed to Masha Y. Ivanova, Center for Children, Youth and Families, University of Vermont Department of Psychiatry, 1 South Prospect St., Burlington, VT 05401; E-mail: Masha.Ivanova@uvm.edu

another society, it is important to determine empirically whether they measure the same constructs in both societies.

Translations of standardized assessment instruments offer valuable opportunities for testing the applicability of these instruments in different societies. A key step in demonstrating the applicability of an assessment instrument involves testing whether patterns of co-occurring problems identified by the instrument in one society fit the patterns identified by the instrument in the other society. Such patterns of problems can be thought of as "syndromes" (i.e., sets of problems that tend to co-occur). To test the degree to which syndromes found in one society fit data from another society, multivariate analyses are performed on data obtained with the instrument for a large sample of children in the new society. If the syndrome structure derived in the instrument's society of origin fits the data obtained in the new society, the original syndrome structure can be applied in the new society. It is important to note that the obtained syndromes may also be shaped by methodological characteristics of the study, including specific analytic methods, recruitment procedures, sample characteristics, and vicissitudes of translations.

The degree to which an instrument's syndrome structure fits a new data set is termed *configural invariance*. Configural invariance is the most basic component of measurement invariance. Measurement invariance refers to the notion that an assessment instrument measures the same psychological constructs in different populations. In addition to including configural invariance, measurement invariance includes metric invariance (i.e., invariance of factor loadings), scalar invariance (i.e., invariance of intercepts), item residual invariance (i.e., invariance of item error variances), factor variance invariance (i.e., invariance of factor variances), factor covariance invariance (i.e., invariance of factor covariances), and factor mean invariance (i.e., invariance of factor means; Vandenberg & Lance, 2000). The components of measurement invariance can be conceptualized as a pyramid, with configural invariance as the base on which the other components rest. Testing configural invariance is the first step in establishing an instrument's measurement invariance (Vandenberg & Lance, 2000).

Factor-Analytic Methods for Deriving and Testing Syndromes

Factor-analytic methods are used to derive syndromes. The first stages usually involve exploratory factor analysis (EFA). EFA is applied to correlations among ratings of items to find patterns of problems that tend to co-occur. After EFA has identified patterns, the patterns can be tested via confirmatory factor analysis (CFA). CFA tests the degree to which a particular model for associations among problems (such as syndromes derived by EFA or by other means) fits a particular data set. If CFA shows good fit between a particular syndrome model and a data set different from the set on which the model was derived, the syndrome model is concluded to have configural invariance in the new data set.

Description of the Present Study

The present study was designed to test the configural invariance of the Teacher's Report Form (TRF) syndrome structure in data sets from 20 societies other than the United States (U.S.) where the syndrome structure was derived. The TRF is part of a multiinformant family of empirically based assessment instruments developed by Achenbach and colleagues (Achenbach & Rescorla, 2001; Achenbach, 1991). It assesses 120 emotional, behavioral, and social problems reported by teachers of students ages 6–18.

The TRF syndromes tested in this study were derived from EFA and CFA of teachers' ratings of combined general population and clinical samples of 4,437 6- to 18-year-old students as detailed by Achenbach and Rescorla (2001). The EFA employed exploratory unweighted least squares (ULS) analyses of polychoric correlations and principal components analyses (PCA) of Pearson correlations. (Polychoric correlations are used with ordinal variables). The CFA employed techniques that were robust to violations of multivariate nor-

mality: Syndromes derived from EFA were fitted on tetrachoric correlations using weighted least squares with standard errors and mean- and variance-adjusted χ^2 test statistic (WLSMV). (Tetrachoric correlations are used with dichotomous variables).

The best-fitting structure for the U.S. data comprised a correlated seven-syndrome model including the Anxious/Depressed, Withdrawn/Depressed, Somatic Complaints, Social Problems, Thought Problems, Rule-Breaking Behavior, and Aggressive Behavior syndromes, plus a hierarchical three-syndrome model for attention problems. The hierarchical model consisted of a general Attention Problems syndrome and two subordinate syndromes of Inattention and Hyperactivity-Impulsivity. In other words, the seven-syndrome and hierarchical Attention Problems models fit the U.S. data better when they were tested separately than when they were tested as part of a single model that combined them. Furthermore, the hierarchical Attention Problems structure was consistent with findings by Dumenci, McConaughy, and Achenbach (2004), who conducted CFA of TRF data obtained for U.S. general population and clinical samples.

The Achenbach and Rescorla (2001) TRF syndromes were highly correlated with previously published TRF syndromes, which did not include a hierarchical substructure for the Attention Problems syndrome (Achenbach, 1991). Before the present study, the 2001 TRF syndrome model had not been tested in non-U.S. societies, but several studies have tested the 1991 U.S. TRF syndromes in non-U.S. samples.

Factor Analyses of the TRF in Other Societies

De Groot, Koot, and Verhulst (1996) tested the 1991 TRF syndrome model using a split-sample design with a clinical sample of 2,442 5- to 18-year-olds evaluated in six Dutch mental health centers. EFA was applied to half of the sample, followed by CFA, which tested both the Dutch and the U.S. EFA models with the other half of the sample. EFA consisted of principal factor analyses with promax rotation, and CFA consisted of ULS of polychoric correlations. The EFA yielded an eight-syndrome model for the Dutch data that was similar to the 1991 U.S. TRF model. The CFA indicated that both the Dutch and U.S. models fit the Dutch data well.

Liu, Kurita, Guo, Tachimori, Ze, and Okawa (2000) performed EFA on TRF ratings for a general population sample of 6- to 11-year-olds in mainland China. The factor derivation sample comprised 454 students whose Total Problems scores were ⩾90th percentile, but who were identified by their teachers as not needing mental health services. PCA with varimax rotation yielded six syndromes, which were designated as Aggressive/Delinquent, Withdrawn/Depressed, Somatic Complaints, Attention Problems, Social Problems, and Thought Problems. The Chinese Aggressive/Delinquent syndrome incorporated most items of the 1991 U.S. Aggressive and Delinquent syndromes, whereas the Chinese Withdrawn/Depressed syndrome incorporated most items of the 1991 U.S. Withdrawn and Anxious/Depressed syndromes. The other four Chinese syndromes also shared most of their items with corresponding 1991 TRF syndromes.

The most extensive test of the 1991 TRF syndrome model in non-U.S. samples was conducted by Hartman et al. (1999). Using CFA, Hartman et al. tested the 1991 model with general population samples from Greece, the Netherlands, Israel, Norway, Portugal, and Turkey. First, the Maximum Likelihood estimation method was applied to Pearson correlations among items. The Root Mean Square Error of Approximation (RMSEA; Browne & Cudeck, 1993), which has been recommended as the best model fit index (Loehlin, 1998), indicated that the 1991 eight-syndrome model fit the data well for all samples. However, other model fit indices suggested that the model did not fit the data. Second, ULS was applied to polychoric correlations among items. The model did not converge for three samples, and all model fit indices suggested poor fit for the remaining three. As Hartman et al. observed, their data violated the assumption of multivariate normality, which is a requirement for the Maximum Likelihood esti-

mation. Because of this violation and different findings with different fit indices and factor-analytic methods, it is difficult to draw firm conclusions from the Hartman et al. results.

In summary, several studies have evaluated the comparability of syndrome structures derived from TRF ratings of emotional and behavioral problems in non-U.S. samples. Except for the Hartman et al. (1999) study, these studies tested the syndrome structure of a translation of the TRF in a single new culture.

Specific Aims of the Present Study

The present study tested whether teachers' ratings of a wide spectrum of students' emotional and behavioral problems would fit the same syndrome structure in very diverse societies. Specifically, we tested the configural invariance of the 2001 TRF syndrome model (Achenbach & Rescorla, 2001) in samples from 20 societies.

This study differed from previous tests of the comparability of syndrome structures derived from teachers' ratings of children's emotional and behavioral problems by (a) testing TRF syndromes in 20 societies, (b) capitalizing on recent advances in CFA methodology, (c) using uniform factor-analytic procedures in each of the 20 societies, and (d) testing the 2001 TRF factor model.

Method

Samples

We used teachers' ratings of problem items from 18 countries, plus Hong Kong and Puerto Rico. Table 1 lists the main reference, sample size, age range, gender distribution, response rate, description of the sampling procedure, and the number of TRFs completed by each teacher for each sample. Note that data from these samples were also used by Rescorla et al. (2007) in their multicultural comparisons of distributions of TRF scale scores.

Sample sizes ranged from 359 for Thailand to 4,857 for China, for a total of 30,030 students ages 6–15. We excluded students older than 15 because many societies have a school-leaving age of 16. Retaining data for students over 15 might have resulted in overrepresentation of students who were especially successful in school.

The gender distribution was about equal in every sample, and the response rate ranged from 72% to 100%. For 12 samples, teachers rated $\leq$2 students in their class, and for 3 samples, teachers rated 3–6 students. For 5 samples (Finland, Lebanon, Poland, Portugal, and Romania), teachers rated all students in their class.

Consistent with standard procedures for analyzing TRF data (Achenbach & Rescorla, 2001), forms on which teachers omitted ratings of more than eight problem items were excluded from analyses. For 15 societies, $\leq$5% of TRFs were excluded for missing $\geq$8 items, but higher percentages were excluded for Iran (7%), Lebanon (13%), Hong Kong (15%), and Romania (22%).

Assessment Instrument

Translations of the TRF were used in all societies, except Jamaica and Australia. A typical translation procedure involved translation of the TRF into the native language by professional translators, followed by back-translation to English to ensure that the translation adequately captured the original meaning. The research team then inspected each translated item to ensure that it captured the meaning of the original item well and would be understood by teachers.

Three samples (Iran, Lithuania, and Romania) were assessed with the 2001 edition of the TRF. The 2001 version replaced three problem items of the 1991 edition with three new items (5. There is very little that he/she enjoys, 28. Breaks school rules, and 99. Smokes, chews, or sniffs tobacco). The remaining samples were assessed with the 1991 edition. To make the results of our analyses comparable among samples, we used the 109 items comprising the eight-syndrome model that were common to the 1991 and 2001 editions. We omitted the three new items, plus the open-ended items (56h. Other physical prob-

Table 1
Samples Used for Confirmatory Factor Analyses of the TRF in 20 Societies

Society	Reference	*N*	Ages	% Male	Response Rate[a]	Sampling Procedure	TRFs per Teacher
Australia	Zubrick et al., 1997	1,697	6–15	49.4	78	Household sample in Western Australia, random sample stratified by region	≤2
China	Liu et al., 2000	4,857	6–15	50.4	92	Household sample in 6 districts in Shandong Province, based on municipal household registers	≤2
Denmark	Bilenberg, 1999	599	6–15	50.4	84	Household sample, regional stratified random general population sample from household registers in Fynen	≤2
Finland	Weintraub, 2004	1,695	6–15	49.2	92	School sample, all children in 134 classes of 8 public schools in southern Finland	Whole class
France	Fombonne & Vermeersch, 1997	493	6–11	48.5	Unknown	Mail survey of employees of a national utility company, 1 child per family	≤2
Greece	Roussos et al., 1999	1,179	6–12	49.3	99	School sample, random sample of 5 children per grade in 6 grades of 42 schools	≤8
Hong Kong	Leung et al., 2006	1,993	6–15	49.7	83	School sample, 89 city schools	≤2
Iran	Minaei, 2005	1,025	6–15	56.9	100	School sample, 10 students in each of 180 schools in Tehran	≤2
Italy	Frigerio et al., 2004	1,022	6–15	49.6	79	School sample, random sample in 3 provinces of Northern Italy, 1 class per grade in 61 schools	≤8
Jamaica	Lambert & Lyubansky, 1999	615	6–15	49.6	87	School sample, random sample in 45 schools in 2 region	≤2
Japan	Kanbayashi 2000	2,559	6–15	50.1	82	School sample, national sample in 72 schools in 10 prefectures	≤8
Lebanon	Bathiche, 2005	1,504	6–11	54.7	95	School sample, 121 classrooms in 18 schools	Whole class
Lithuania	Zukauskiene & Kajokiene, 2004	2,601	6–15	49.5	84	School sample, national sample of 2 classes per grade in 153 schools	≤2
Netherlands	Verhulst, Van der Ende, Ferdinand, & Kasius, 1997	1,239	6–15	51.6	83	Household sample, national probability sample of Dutch children	≤2
Poland	Wolanczyk, 2003	2,133	6–15	48.2	99	School sample, national sample of 132 schools, one class per grade	Whole class
Portugal	Fonseca et al., 1995	1,373	6–15	52.2	91	School sample, regional stratified sample in Coimbra County	Whole class
Puerto Rico	Achenbach et al., 1990	565	6–15	51.0	93	Household sample, island wide sample in 210 population clusters	≤2
Romania	Domuta, 2004	922	6–11	49.3	80	School sample, four towns	Whole class
Thailand	Weisz et al., 1989	359	6–11	50.1	92	School sample, national random sample in 38 schools	≤2
Turkey	Erol & Simsek, 1997	1,600	6–15	49.0	88	Household sample, national general population	≤2
		Total: 30,030					

Note. TRF = Teacher's Report Form.
[a]Percentage of targeted students for whom TRFs were completed.

lems without known medical cause and 113. Other problems).

Tested Models

Figures 1 and 2 illustrate the tested models. The first model included the seven correlated factors of Anxious/Depressed, Withdrawn/Depressed, Somatic Complaints, Social Problems, Thought Problems, Rule-Breaking Behavior, and Aggressive Behavior. The second model was a hierarchical three-factor structure comprising the general Attention Problems factor and two subordinate specific factors of Inattention and Hyperactivity-Impulsivity (Achenbach & Rescorla, 2001; Dumenci et al., 2004). For the hierarchical model, factor covariances were fixed at 0.

Data Analysis

Teachers rated each item for each child on a 3-point scale (0 = *not true [as far as you know]*, 1 = *somewhat or sometimes true*, and 2 = *very true or often true*), based on the preceding 2 months. Following Achenbach and Rescorla's (2001) procedures, we dichotomized the data by converting item scores to 0 versus 1 or 2 in order to use nonparametric tetrachoric correlations. (This was done to make the data appropriate for analyses that were robust to violations of multivariate normality.) To account for the non-normal distributions of the item ratings, we used WLSMV via Mplus 3.0 (Muthén & Muthén, 2004). Mplus omitted items 101 (Truancy) and 105 (Uses alcohol or drugs for nonmedical purposes) from the CFA of the French sample because of insufficient variance.

We used the RMSEA as the primary model fit index because it has been identified as the best performing index for WLSMV (Yu & Muthén, 2002). The RMSEA is an absolute fit index that evaluates the degree of misfit per degrees-of-freedom (Hu & Bentler, 1999). To follow the convention of using multiple fit indices, we also computed the Comparative Fit Index (CFI; Bentler, 1990) and the Tucker-Lewis Index (TLI; Tucker & Lewis, 1973). The CFI and TLI are incremental fit indices that evaluate the proportional improvement in

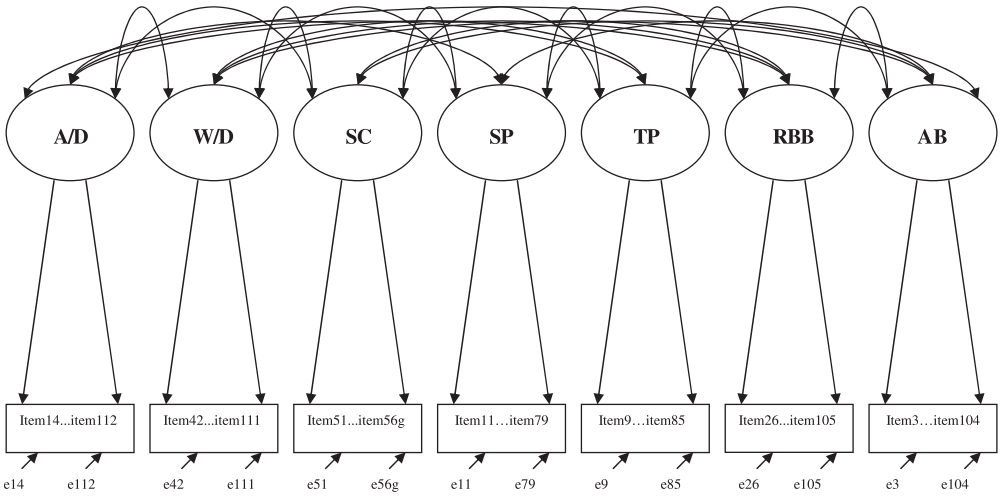

Figure 1. The seven-factor Teacher's Report Form model tested in the study: A/D = Anxious/Depressed; W/D = Withdrawn/Depressed; SC = Somatic Complaints; SP = Social Problems; TP = Thought Problems; RBB = Rule-Breaking Behavior; AB = Aggressive Behavior. Items 101 and 105 were included for all societies, except France.

model fit by comparing the tested model with a more restricted baseline model (Hu & Bentler, 1999). Because it is still unknown whether the CFI and TLI are appropriate for use with categorical data, we considered their results secondary to the results of the RMSEA. The current guidelines for the evaluation of model fit range from <.05 (Yu

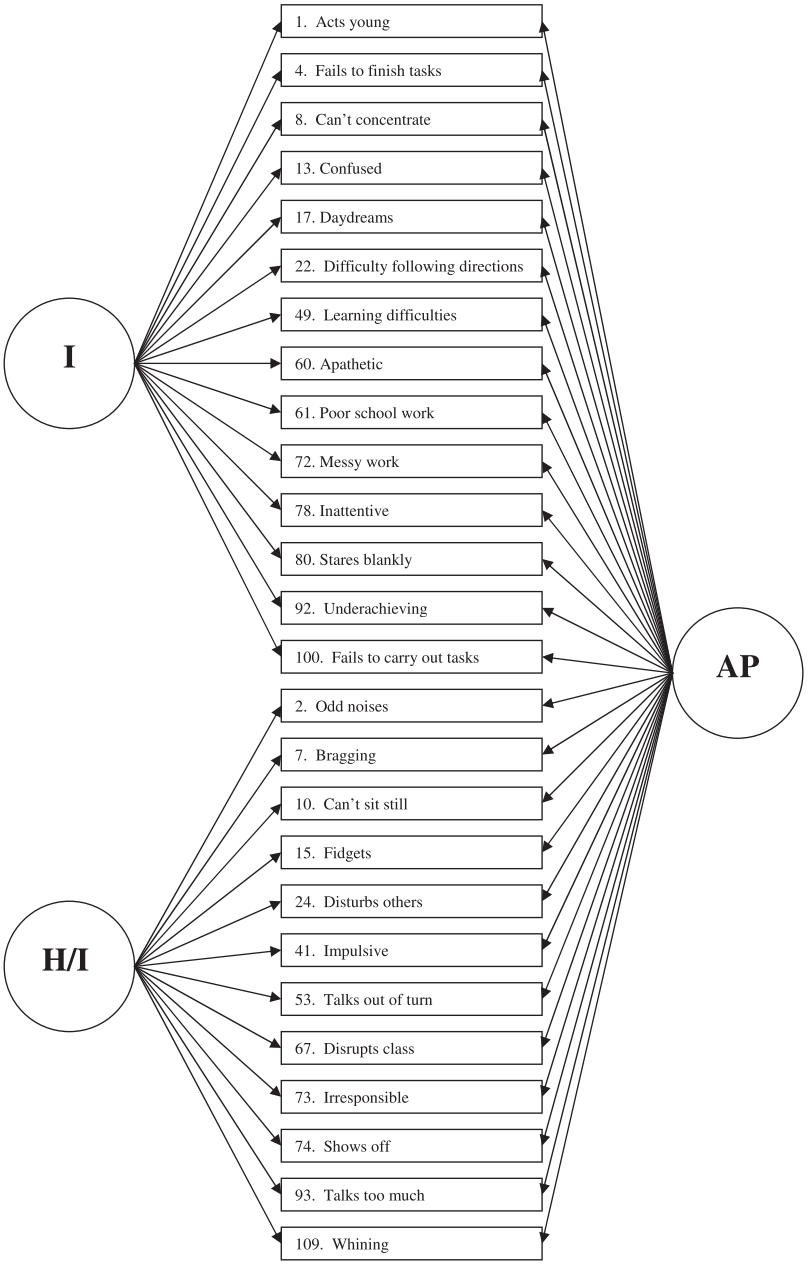

Figure 2. The hierarchical three-factor Teacher's Report Form model tested in the study: AP = Attention Problems; I = Inattention; H/I = Hyperactivity-Impulsivity. Items 101 and 105 were included for all societies, except France.

Table 2
Results of Confirmatory Factor Analyses for 20 Societies

		Seven-Syndrome Model					Three-Syndrome Model			
Society	*N*	RMSEA	CFI	TLI	Median Item Loading	Median Factor Correlation	RMSEA	CFI	TLI	Median Item Loading
Australia	1,697	.048	.889	.943	.80	.67	.052	.966	.992	.59
China	4,857	.037	.827	.921	.69	.76	.037	.950	.981	.52
Denmark	599	.063	.933	.947	.78	.66	.050	.979	.993	.61
Finland	1,695	.050	.921	.941	.81	.66	.047	.971	.991	.57
France	493	.065	.826	.836	.72	.50	.056	.964	.982	.59
Greece	1,179	.071	.860	.887	.76	.60	.061	.943	.985	.58
Hong Kong	1,993	.052	.732	.950	.81	.73	.050	.962	.991	.60
Iran	1,025	.063	.828	.940	.78	.77	.048	.964	.991	.61
Italy	1,022	.060	.814	.836	.72	.52	.060	.951	.983	.62
Jamaica	615	.058	.866	.905	.71	.62	.056	.954	.982	.54
Japan	2,559	.034	.885	.962	.80	.78	.040	.961	.991	.59
Lebanon	1,504	.063	.840	.883	.74	.56	.053	.956	.987	.58
Lithuania	2,601	.070	.740	.875	.72	.62	.047	.958	.986	.50
Netherlands	1,239	.059	.873	.910	.77	.64	.059	.943	.981	.52
Poland	2,133	.076	.811	.880	.73	.56	.069	.942	.982	.51
Portugal	1,373	.068	.821	.894	.77	.62	.057	.963	.988	.59
Puerto Rico	565	.064	.884	.920	.73	.60	.064	.968	.990	.58
Romania	922	.069	.826	.887	.72	.67	.061	.943	.982	.58
Thailand	359	.055	.919	.943	.71	.73	.048	.972	.987	.53
Turkey	1,600	.064	.821	.885	.71	.61	.057	.948	.986	.60

Note. RMSEA = Root Mean Square Error of Approximation; CFI = Comparative Fit Index; TLI = Tucker–Lewis Index. The seven-syndrome model is the correlated model comprising the Anxious/Depressed, Withdrawn/Depressed, Somatic Complaints, Social Problems, Thought Problems, Rule-Breaking Behavior, and Aggressive Behavior syndromes. The three-syndrome model is the hierarchical model comprising the general Attention Problems syndrome and the subordinate Inattention and Hyperactivity-Impulsivity syndromes. For the three-syndrome model, factor correlations were set to 0. For France, the seven-syndrome model excluded items 101 and 105 because of insufficient item variance.

& Muthén, 2002) to <.08 (Browne & Cudeck, 1993) for the RMSEA, and from >.90 (Browne & Cudeck, 1993) to >.95 (Hu & Bentler, 1999) for the CFI and TLI. However, Marsh, Hau, and Wen (2004) argued that Hu and Bentler's criteria were too stringent because they significantly increased the risk of rejecting properly defined complex models. Given the complexity of our model, we used Browne and Cudeck's (1993) less stringent criteria of <.08 for the RMSEA and >.90 for the CFI and TLI as indicating good model fit. For the CFI and TLI, we considered values between .80 and .90 to indicate acceptable fit.

Results

Overall Model Fit

As Table 2 shows, the models converged for every sample. For the seven-syndrome model, the RMSEA ranged from .034 (Japan) to .076 (Poland), indicating good fit for every society. To give a better sense of the distribution of RMSEA values across societies, RMSEA equaled .053, .063, and .067 at the 25th, 50th, and 75th percentiles, respectively. The CFI ranged from .732 (Hong Kong) to .933 (Denmark), indicating an acceptable to good model fit in each society,

except Hong Kong and Lithuania (CFI = .740). The TLI ranged from .836 (France) to .962 (Japan), indicating an acceptable to good model fit for each of the 20 societies. For the hierarchical Attention Problems model, the RMSEA ranged from .037 (China) to .069 (Poland), indicating good model fit for each of the 20 societies. RMSEA values equaled .048, .055, and .060 at the 25th, 50th, and 75th percentiles, respectively. The CFI ranged from .942 to .979, and the TLI ranged from .981 to .993, indicating good model fit for each of the 20 societies for both indexes.

Items With Negative Error Variance

As Table 3 shows, for the seven-syndrome model, 11 societies had no items with negative error variances. For the remaining societies, the following numbers of items had negative error variances: 1 item for Australia, Greece, Italy, the Netherlands, and Romania; 2 items for Denmark, Finland, and France; and 3 items for Puerto Rico. For the hierarchical Attention Problems model, all items had positive error variances for 19 societies, but 1 item had a negative error variance for Greece. Thus, only 14 out of 3,740 (.0037) estimated parameters for the seven-syndrome model and 1 out of 1,620 (.0006) estimated parameters for the hierarchical Attention Problems model were outside of the admissable parameter space.

Item Loadings

As shown in Table 3, for the seven-syndrome model, all 83 items loaded significantly on their predicted factors for each of 16 societies. However, 1 item for France, Jamaica, and Romania, and 2 items for the Netherlands failed to load on their predicted factors. Also, as shown in Table 3, for the hierarchical Attention Problems model, all 26 items loaded significantly on their predicted factors for Greece, Lebanon, and Turkey. In the other societies, the following numbers of items failed to load significantly on their predicted factors: 1 item for Australia, Hong Kong, Japan, Portugal, and Romania; 2 items for Finland, France, Iran, Italy, the Netherlands, and Poland; 3 items for China, Jamaica, Lithuania, and Puerto Rico; 5 items for Thailand; and 13 items for Denmark. For Denmark, the 13 nonsignificant loadings included all 12 items of the Hyperactivity-Impulsivity syndrome.

Median Item Loadings and Factor Correlations

Table 2 presents median item loadings for the seven- and three-syndrome models for each society. For the seven-syndrome model, the median item factor loadings ranged from .69 (China) to .81 (Finland and Hong Kong), with an overall median of .74. For the three-syndrome model, the median item factor loadings ranged from .50 (Lithuania) to .62 (Italy), with an overall median of .58.

Table 4 presents the median of each item's loading from the CFA done in each of the 20 societies. For the seven-syndrome model, the medians of the item loadings ranged from .37 (32. Feels he/she has to be perfect) to .91 (54. Overtired without good reason and 103. Unhappy, sad, or depressed), with an overall median of .76. For the three-syndrome model, the medians of the item loadings ranged from .07 (109. Whining on the Hyperactive-Impulsive factor) to .90 (73. Behaves irresponsibly on the Attention Problems factor), with an overall median of .56. The median of the item loadings within syndromes from the 20 societies ranged from .47 (Inattention) to .82 (Aggressive Behavior).

Finally, Table 2 presents median factor correlations in each society for the seven-syndrome model. They ranged from .50 (France) to .78 (Japan), with the overall median of .63. As illustrated in Figure 2, factor correlations were not free parameters in the three-factor model—hence, they were fixed at 0.

Discussion

Results of this study indicated that the seven-syndrome model and the hierarchical Attention Problems model that were the best-fitting models for U.S. data also fit the data in samples from each of the 20 societies. For both tested models, the RMSEA was $<.08$ for

Table 3
Results of Confirmatory Factor Analyses for 20 Societies: Items With Negative Error Variances and Nonsignificant Factor Loadings

	Seven-Syndrome Model		Three-Syndrome Model	
Society	Empirically Underidentified Items (i.e., Items With Negative Error Variance)[a]	Items With Nonsignificat Factor Loadings[a]	Empirically Underidentified Items (i.e., Items With Negative Error Variance)[a]	Items With Nonsignificant Factor Loadings[a]
Australia	103			15(H/I)
China				22(I), 53(H/I), 109(H/I)
Denmark	25, 103			2(H/I), 7(H/I), 10(H/I), 15(H/I), 24(H/I), 41(H/I), 53(H/I), 67(H/I), 73(H/I), 74(H/I), 78(I), 93(H/I), 109(H/I)
Finland	54, 103			2(H/I), 109(H/I)
France	40, 54	98		73(H/I), 80(AP)
Greece	45		73	
Hong Kong				73(H/I)
Iran				41(H/I), 73(H/I)
Italy	51			73(H/I), 109(H/I)
Jamaica		91		7(AP), 72(I), 100(I)
Japan				109(H/I),
Lebanon				
Lithuania				49(AP), 73(H/I), 109(H/I),
Netherlands	103	56d, 56e		73(H/I), 109(H/I)
Poland				13(AP), 109(H/I)
Portugal				109(H/I)
Puerto Rico	54, 91, 103			60(I), 72(I), 73(H/I)
Romania	56g	71		109(H/I)
Thailand				1(I), 24(H/I), 41(H/I), 73(H/I), 78(I)
Turkey				

Note. The seven-syndrome model is the correlated model comprising the Anxious/Depressed, Withdrawn/Depressed, Somatic Complaints, Social Problems, Thought Problems, Rule-Breaking Behavior, and Aggressive Behavior syndromes. The three-syndrome model is the hierarchical model comprising the general Attention Problems (AP) syndrome and the subordinate Inattention (I) and Hyperactivity-Impulsivity (H/I) syndromes.
[a]Each number is the number that an item bears on the TRF.

each society, indicating acceptable model fit. The results of the CFI and TLI were consistent with the RMSEA findings, with the exception of the CFI values for Hong Kong and Lithuania for the seven-syndrome model. The very few estimated parameters that were outside of the permissible parameter space ($<.004$ for the seven-syndrome model and $<.001$ for the hierarchical Attention Problems model) may have reflected sampling fluctuations. Across

Table 4
Median Item Loadings on TRF Syndromes Across 20 Societies

Syndromes and Items	Median Loading	Syndromes and Items	Median Loading
Anxious/Depressed	(.69)	Thought Problems	(.72)
14. Cries a lot	.65	9. Can't get mind off thoughts	.72
29. Fears	.51	18. Harms self	.71
30. Fears school	.67	40. Hears things	.78
31. Fears doing bad	.52	46. Twitching	.72
32. Must be perfect	.37	58. Picks skin	.65
33. Feels unloved	.85	66. Repeats acts	.71
35. Feels worthless	.77	70. Sees things	.68
45. Nervous	.81	83. Stores things	.63
50. Anxious	.72	84. Strange behavior	.82
52. Feels guilty	.70	85. Strange ideas	.77
71. Self-conscious	.56	Rule-Breaking Behavior	(.69)
81. Feels hurt	.70	26. No guilt after misbehavior	.82
91. Talks about suicide	.74	39. Hangs around peers who get in trouble	.80
106. Anxious to please	.54	43. Lying or cheating	.82
108. Afraid of mistakes	.56	63. Prefers older children	.51
112. Worries	.70	82. Steals	.73
Withdrawn/Depressed	(.77)	90. Swears	.86
42. Rather be alone	.77	96. Preoccupied with sex	.65
65. Won't talk	.77	98. Tardy	.60
69. Secretive	.72	101. Truant	.65
75. Shy	.58	105. Uses drugs or alcohol	.60
102. Underactive	.73	Aggressive Behavior	(.82)
103. Sad	.91	3. Argues	.76
111. Withdrawn	.81	6. Defiant	.82
Somatic Complaints	(.77)	16. Cruel	.84
51. Dizzy	.81	19. Demands attention	.72
54. Overtired	.91	20. Destroys own things	.78
56a. Aches	.77	21. Destroys others' things	.83
56b. Headaches	.77	23. Disobedient at school	.83
56c. Nausea	.85	37. Gets in many fights	.85
56d. Eye problems	.46	57. Attacks	.85
56e. Skin problems	.50	68. Screams	.81
56f. Stomachaches	.76	76. Explosive	.86
56g. Vomiting	.77	77. Low frustration tolerance	.78
Social Problems	(.64)	86. Stubborn	.81
11. Dependent	.53	87. Sudden mood changes	.81
12. Lonely	.64	88. Sulks	.74
25. Doesn't get along	.84	89. Suspicious	.78
27. Jealous	.71	94. Teases	.83
34. Others out to get him/her	.79	95. Temper tantrums	.88
36. Accident-prone	.64	97. Threatens	.89
38. Teased	.73	104. Loud	.79
48. Unliked	.79	Attention Problems	(.68)
62. Clumsy	.64	1. Acts young	.66
64. Prefers younger kids	.49	2. Odd noises	.70
79. Speech problems	.43	4. Fails to finish tasks	.71

(Table 4 continues)

(Table 4 continued)

Syndromes and Items	Median Loading	Syndromes and Items	Median Loading
7. Bragging	.46	4. Fails to finish tasks	.46
8. Can't concentrate	.78	8. Can't concentrate	.38
10. Can't sit still	.70	13. Confused	.53
13. Confused	.49	17. Daydreams	.56
15. Fidgets	.73	22. Difficulty following directions	.40
17. Daydreams	.45	49. Learning difficulties	.70
22. Difficulty following directions	.76	60. Apathetic	.64
24. Disturbs others	.77	61. Poor school work	.69
41. Impulsive	.77	72. Messy work	.29
49. Learning difficulties	.53	78. Inattentive	.33
53. Talks out of turn	.60	80. Stares blankly	.63
60. Apathetic	.53	92. Underachieving	.47
61. Poor school work	.59	100. Fails to carry out tasks	.47
67. Disrupts class	.76	Hyperactivity-Impulsivity	(.46)
72. Messy work	.78	2. Odd noises	.34
73. Irresponsible	.90	7. Bragging	.55
74. Shows off	.65	10. Can't sit still	.46
78. Inattentive	.82	15. Fidgets	.38
80. Stares blankly	.44	24. Disturbs others	.46
92. Underachieving	.64	41. Impulsive	.36
93. Talks too much	.57	53. Talks out of turn	.62
100. Fails to carry out tasks	.74	67. Disrupts class	.53
109. Whining	.54	73. Irresponsible	.09
Inattention	(.47)	74. Shows off	.54
1. Acts young	.26	93. Talks too much	.62
		109. Whining	.07

Note. For items 11 and 105, median item loadings were calculated across all societies, except France. Parentheses show median of the median loadings on each syndrome.

all 20 societies, the median loadings of items on their predicted factors were .74 for the seven-syndrome model and .56 for the hierarchical Attention Problems model. The findings that similar syndrome structures fit data obtained with the TRF items in each of the tested societies increase confidence in the similarity of constructs captured by scores on TRF syndromes in the U.S. and each of the 20 other societies.

The degree to which the TRF syndrome structures fit data from 20 societies is impressive, given the multitude of factors that could have contributed to differences among samples. Our samples represented diverse world regions, including Asia; Australia; the Caribbean; eastern, western, southern, and northern Europe; and the Middle East. The samples came from societies that vary tremendously in their political, educational, and health systems, as well as in child-rearing practices and religion. Methodological differences among the samples, such as variations in recruitment procedures, TRF translations, and the number of students rated by each teacher, could have affected the results. In addition, sample characteristics, such as response rate and the students' age, could also have caused differences in findings.

The results of this study are consistent with findings for parent and youth self-ratings of emotional and behavioral problems. Ivanova et al. (2007, in press) tested the fit of

Table 5
Means and Standard Deviations (in Parentheses) for TRF Syndromes in 20 Societies

	TRF Syndromes									
Society	A/D	W/D	SC	SP	TP	RBB	AB	AP	I	H/I
Australia	2.29 (3.2)	1.13 (1.9)	0.51 (1.2)	1.25 (2.2)	0.57 (1.3)	0.86 (1.8)	2.34 (4.8)	7.41 (9.5)	4.54 (6.0)	2.87 (4.5)
China	1.67 (2.5)	1.25 (2.0)	0.49 (1.2)	1.32 (1.9)	0.31 (0.9)	0.99 (1.6)	2.31 (3.5)	5.67 (6.7)	3.65 (4.4)	2.02 (2.9)
Denmark	3.76 (4.4)	1.41 (2.1)	0.62 (1.6)	2.20 (3.3)	0.67 (1.7)	1.26 (2.3)	4.67 (7.3)	7.94 (10.3)	4.75 (6.3)	3.19 (4.9)
Finland	1.12 (2.1)	0.90 (1.6)	0.30 (0.9)	0.89 (1.9)	0.20 (0.8)	0.63 (1.7)	2.08 (4.8)	4.31 (7.1)	2.45 (4.1)	1.86 (3.7)
France	3.74 (3.7)	2.05 (2.2)	0.31 (0.9)	1.06 (1.7)	0.33 (1.0)	0.66 (1.2)	2.59 (4.0)	5.30 (6.6)	2.97 (4.3)	2.33 (3.6)
Greece	4.58 (4.0)	1.70 (2.3)	0.27 (0.7)	1.81 (2.5)	0.35 (1.0)	1.10 (1.7)	3.26 (5.6)	7.57 (9.0)	4.37 (6.0)	3.19 (4.4)
Hong Kong	2.87 (3.5)	1.71 (2.3)	0.33 (1.1)	1.77 (2.6)	0.49 (1.3)	1.03 (2.0)	2.84 (5.1)	7.66 (9.0)	5.20 (5.9)	2.46 (4.0)
Iran	4.92 (4.5)	2.10 (2.8)	0.75 (1.6)	2.39 (3.0)	1.01 (2.0)	1.46 (2.3)	4.75 (6.4)	8.57 (9.9)	4.52 (5.6)	4.06 (5.1)
Italy	2.81 (3.4)	1.70 (2.2)	0.37 (1.0)	1.35 (2.1)	0.27 (0.8)	0.92 (1.6)	2.14 (3.8)	7.46 (8.4)	4.89 (5.7)	2.57 (4.0)
Jamaica	4.06 (3.8)	2.52 (2.9)	0.76 (1.5)	1.96 (2.6)	0.89 (1.6)	2.18 (2.8)	4.76 (6.3)	12.35 (10.3)	7.56 (6.8)	4.79 (5.1)
Japan	1.10 (2.1)	0.45 (1.1)	0.20 (0.8)	0.89 (2.0)	0.19 (0.8)	0.40 (1.2)	1.40 (3.7)	3.43 (6.5)	2.16 (4.0)	1.27 (2.9)
Lebanon	3.59 (3.8)	1.60 (2.4)	0.35 (1.1)	1.71 (2.5)	0.56 (1.3)	0.94 (1.7)	3.37 (5.4)	8.43 (9.2)	5.00 (6.1)	3.43 (4.5)
Lithuania	4.34 (3.6)	1.89 (2.2)	1.15 (1.8)	2.62 (2.6)	0.72 (1.4)	1.80 (2.4)	4.47 (6.0)	8.97 (8.2)	5.23 (4.9)	3.74 (4.3)
Netherlands	3.12 (3.7)	1.73 (2.2)	0.42 (1.1)	1.84 (2.8)	0.55 (1.2)	0.83 (1.6)	3.05 (5.3)	7.86 (8.5)	4.72 (5.2)	3.14 (4.3)
Poland	4.26 (3.9)	2.54 (2.7)	0.85 (1.5)	2.00 (2.6)	0.46 (1.1)	1.76 (2.8)	4.69 (6.6)	9.20 (9.0)	5.66 (5.5)	3.54 (4.8)
Portugal	3.57 (3.4)	2.03 (2.4)	0.47 (1.2)	1.51 (2.3)	0.41 (1.3)	1.32 (2.2)	3.17 (5.5)	8.89 (9.1)	5.66 (6.0)	3.23 (4.3)
Puerto Rico	4.19 (3.9)	3.98 (3.4)	0.90 (1.5)	2.09 (2.7)	0.76 (1.4)	1.47 (2.4)	4.80 (6.8)	12.51 (11.6)	8.31 (8.0)	4.20 (5.2)
Romania	4.35 (4.0)	1.57 (2.1)	0.54 (1.4)	2.60 (2.9)	0.93 (1.6)	1.52 (1.9)	4.89 (6.6)	8.21 (9.1)	4.47 (5.5)	3.74 (4.9)
Thailand	4.89 (4.0)	2.04 (2.2)	1.27 (2.2)	2.43 (2.4)	1.28 (2.0)	1.47 (2.1)	4.77 (5.6)	11.05 (9.2)	6.49 (5.6)	4.56 (4.4)
Turkey	4.60 (4.0)	1.97 (2.6)	0.56 (1.3)	1.81 (2.6)	0.69 (1.4)	0.94 (1.9)	3.81 (5.3)	8.00 (9.2)	4.50 (6.1)	3.50 (4.7)

Note. Raw scores were used. A/D = Anxious/Depressed; W/D = Withdrawn/Depressed; SC = Somatic Complaints; SP = Social Problems; TP = Thought Problems; RBB = Rule-Breaking Behavior; AB = Aggressive Behavior; AP = Attention Problems; I = Inattention; H/I = Hyperactivity-Impulsivity.

the eight-syndrome model scored from the Child Behavior Checklist (CBCL) and the Youth Self-Report (YSR; Achenbach & Rescorla, 2001). The CBCL and YSR are parent- and self-report counterparts to the TRF that share an eight-syndrome structure quite similar to the TRF syndromes. The eight-syndrome structure comprises the seven syndromes of the TRF seven-syndrome model, plus the general Attention Problems syndrome without the hierarchical substructure. For the CBCL, Ivanova et al. (2007) tested the eight-syndrome model in each of 30 societies, which included all societies tested in the present study, except Lebanon. The model converged in each of the 30 samples, and all 30 RMSEA values indicated good model fit. For the YSR, the syndrome model was tested in each of 23 societies, which included all societies tested in the present study, except China, France, Italy, Lebanon, Portugal, and Thailand. The model converged in each of the 23 societies, and the RMSEA values indicated good model fit in each society (Ivanova et al., in press).

Rescorla et al. (2007) compared TRF syndrome scores for students in the 20 samples used in this study, plus a U.S. general population sample. For the eight syndromes and for scales oriented to the *Diagnostic and Statistical Manual*, the effect sizes for society were in the small to medium range, using Cohen's (1988) criteria. These findings indicated that teachers' reports of students' emotional and behavioral problems are fairly similar among very different societies. However, because significant effect sizes for society were found, there may also be important variations in syndrome scores between some societies. Table 5 presents means and standard deviations of TRF syndrome scores obtained in the present study.

Limitations and Implications

To serve diverse student bodies, school psychologists need psychometrically sound standardized instruments to assess students from different backgrounds. The results of the present study and the Ivanova et al. (2007, in press) studies provide preliminary evidence of the similarity of syndromes measured by the TRF, CBCL, and YSR in very different societies. However, these results should not be interpreted to imply that the TRF, CBCL, and YSR measure the same psychological constructs in all the tested societies. Configural invariance is only the most basic element of measurement invariance. To conclude that an assessment instrument measures the same construct in different societies, it is necessary to formally test all components of measurement invariance. To account for the non-normal distribution of our data, we used the WLSMV estimator in the present study and the Ivanova et al. (2007, in press) studies. The WLSMV is a recently developed advanced estimator that is robust to violations of multivariate normality. Because the WLSMV is so computationally intensive, it is not now feasible to use WLSMV to test components of measurement invariance other than configural invariance.

Because teachers rated students in their own societies, it remains to be seen whether TRF syndromes would fit teachers' ratings of students from societies other than their own. An example of this situation that is becoming increasingly relevant in U.S. schools would be when a U.S. teacher completes a TRF for an immigrant student. Another limitation of the present study is that although the TRF assesses a broad range of emotional and behavioral problems, other important problems might also be assessed. Some of the other important problems might be specific to certain societies. For example, we found that although the hierarchical Attention Problems model fit the Danish data extremely well (RMSEA = .050, CFI = .979, TLI = .993), the factor loadings for items comprising the Hyperactivity-Impulsivity syndrome did not reach statistical significance. This finding may reflect sampling fluctuations, but it is also possible that the translated items comprising the Hyperactivity-Impulsivity factor were not as relevant for Danish teachers as for teachers in other societies. Furthermore, although the samples used in this study represent a wide range of societies, we cannot assume that the TRF syndrome model would fit teachers' ratings in societies that were not represented in this study. The fit of the TRF syndrome model to data obtained in other societies should be tested in future research.

Despite these possible limitations, the results of the present study and the Ivanova et al. (2007, in press) CFA studies of the CBCL and YSR support the tested syndromes as templates for conceptualizing psychological difficulties in diverse societies. The findings of these studies can guide research and practice related to helping students from diverse societies succeed in school. They can also guide multicultural collaboration among school psychologists and other professionals for purposes of training, service, and research.

References

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

Achenbach, T. M., Bird, H. R., Canino, G., Phares, V., Gould, M. S., & 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., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms & Profiles.* Burlington: University of Vermont, Research Center for Children, Youth, and Families.

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

Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin, 107,* 238–246.

Bilenberg, N. (1999). The Child Behavior Checklist (CBCL) and related material: Standardization and validation in Danish population based and clinically based samples. *Acta Psychiatrica Scandinavica, 100 (Suppl. 398),* 2–52.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park: Sage.

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

De Groot, A., Koot, H. M., & Verhulst, F. C. (1996). Cross-cultural generalizability of the Youth Self-Report and Teacher's Report Form cross-informant syndromes. *Journal of Abnormal Child Psychology, 24,* 651–664.

Domuta, A. (2004). *Evaluarea sindromului ADHD la varsta prescolara.* [ADHD assessment in preschoolers.] Unpublished doctoral dissertation, Babes-Bolyai University, 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/emotional problems and competencies in Turkish children. In N. Erol, C. Kilic, M. Ulusoy, M. Kececi, & Z. Simsek (Eds.), *Mental health profiles of Turkey: A preliminary report* (pp. 12–33). Ankara: Aydogdu.

Fombonne, E., & Vermeersch, S. (1997). Children from the GAZEL cohort: I—Prevalence and patterns of service utilization for emotional and behavioural symptoms. *Revue d'Épidémiologie et de Santé Publique, 45,* 29–40.

Fonseca, A. C., Simoes, A., Rebelo, J. A., Ferriera, 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.

Hartman, C. A., Hox, J., Auerbach, J., Erol, N., Fonseca, A. C., Mellenbergh, G. J., et al. (1999). Syndrome dimensions of the Child Behavior Checklist and Teacher Report Form: A critical empirical evaluation. *Journal of Child Psychology and Psychiatry, 40,* 1095–1116.

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

Ivanova, M. Y., Achenbach, T. M., Dumenci, L., Rescorla, L. A., Almqvist, F., Bilenberg, N., et al. (2007). Testing the 8-syndrome structure of the Child Behavior Checklist in 30 societies. *Journal of Clinical Child and Adolescent Psychology, 36,* 1–13.

Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Dumenci, L., Almqvist, F., Bilenberg, N., et al. (in press). The generalizability of the Youth Self-Report syndrome structure in 23 societies. *Journal of Consulting and Clinical Psychology.*

Kanbayashi, Y. (2000). [Teacher's Report Form data]. Unpublished raw data.

Lambert, M. C., & Lyubansky, M. (1999). Behavior and emotional problems among Jamaican children and adolescents: An epidemiological survey of parent, teacher, and self-reports for ages 6–18 years. *International Journal of Intercultural Relations, 23,* 727–751.

Leung, P. W., 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 the CBCL, TRF, and YSR. *Journal of Child Psychology and Psychiatry, 47,* 970–973.

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

Loehlin, J. C. (1998). *Latent variable models: An introduction to factor, path, and structural analysis.* (3rd ed.). Mahwah: Erlbaum.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11,* 320–341.

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

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

Rescorla, L. A., Achenbach, T. M., Ginzburg, S., Ivanova, M. Y., Dumenci, L., Almqvist, F., et al. (2007). Consis-

tency of teacher-reported problems for students in 21 countries. *School Psychology Review*, *36,* 91–110.

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.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38,* 1–10.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3,* 4–70.

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.

Weintraub, S. (2004) [Teacher's Report Form data]. Unpublished raw data.

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

Wolanczyk, T. (2003). *Zaburzenia emocjonalne i behawioralne u dzieci i mlodziezy w Polsce* [Emotional and behavioral problems in children and adolescents in Poland.] Warszaw: Akademia Medyczna w Warszawie.

Yu, C. Y., & Muthén, B. O. (2002). *Evaluation of model fit indices for latent variable models with categorical and continuous outcomes* (Technical Report). Los Angeles: UCLA, Graduate School of Education and Information Studies.

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, Western Australia: Australian Bureau of Statistics and the TVW Telethon Institute for Child Health Research.

Zukauskiene, R., & Kajokiene, I. (2004, October). *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.

Date Received: December 13, 2005
Date Accepted: May 8, 2007
Action Editor: Ruth Ervin ■

Masha Y. Ivanova is Research Assistant Professor of Psychiatry at the University of Vermont. She received her PhD at the State University of New York at Albany and completed a predoctoral clinical internship at the University of Vermont Center for Children, Youth, and Families. 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.

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.

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.

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 four authors.