



Published in final edited form as:

Assessment. 2013 October ; 20(5): 642–655. doi:10.1177/1073191113498114.

Measurement Invariance of Internalizing and Externalizing Behavioral Syndrome Factors in a Non-Western Sample

Lisa M. Yarnell¹, Marsha N. Sargeant¹, Carol A. Prescott¹, Jacqueline L. Tilley¹, Jo Ann M. Farver¹, Sarnoff A. Mednick¹, Peter H. Venables², Adrian Raine³, and Susan E. Luczak^{1,4}

¹University of Southern California, Los Angeles, CA, USA

²University of York, York, UK

³University of Pennsylvania, Philadelphia, PA, USA

⁴University of California, San Diego, CA, USA

Abstract

This study examined the measurement structure of Child Behavior Checklist internalizing and externalizing syndrome scales in 1,146 eleven-year-old children from a birth cohort in Mauritius. We tested for measurement invariance at configural, metric, and scalar levels by gender and religioethnicity (Creole, Hindu, Muslim). A pared-down model representing five primary factors and two secondary factors met all three forms of invariance, supporting the validity of their use for group comparisons among Mauritian children. As rated by their parents, girls were higher than boys on Somatic Complaints and lower on Aggressive Behavior, Attention Problems, and Externalizing. Creoles were higher than Muslims and Hindus on all seven factors. Hindus were higher than Muslims on Somatic Complaints and lower on Aggressive Behavior. To our knowledge, this is the first study to demonstrate strict invariance of a Child Behavior Checklist-based internalizing and externalizing factor structure among subgroups within a society.

Keywords

Joint Child Health Project; Child Behavior Checklist; gender; religion; ethnicity; Indian; African; Mauritius

Knowledge of child psychopathology has typically relied on research conducted in Western cultures. Examining child behavior syndromes across cultures, and among subgroups within a culture, provides a broader understanding of the stability and variability of syndromes and can inform future work on taxonomy building, prevalence rates, and clinical evaluation and treatment (see Weisz, Weiss, Suwanlert, & Chaiyasit, 2003). Using an assessment instrument created in one culture with individuals from another culture, however, is appropriate only to the extent that the instrument measures the same concepts or constructs across these cultures. This is also true for subgroups within a culture, such as gender and ethnic groups, that may vary in their manifestation of psychopathology (Eaton et al., 2011; McLaughlin, Hilt, & Nolen-Hoeksema, 2007). Otherwise, differences found between groups may reflect measurement bias rather than differences at the construct level.

© The Author(s) 2013

Corresponding Author: Susan E. Luczak, Department of Psychology, University of Southern California, SGM 501, Los Angeles, CA 90089-1061, USA. luczak@usc.edu.

Declaration of Conflicting Interests

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

The purpose of this study is twofold: (a) to examine how well the constructs of internalizing and externalizing that have been derived from Western samples apply to a non-Western sample and (b) to adapt the Western structure to create a factor structure that yields equivalent constructs across subgroups (e.g., gender, ethnic groups) within a non-Western sample to make valid comparisons across groups. To do this, we first employ a “forced etic” approach (Berry, 1989), in which we apply the current U.S.-derived behavioral syndrome factor structure of the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001) to a population cohort of children from the African country of Mauritius. After finding the fit of the Western structure to be generally promising, we then proceed to an “emic” approach where we refine the factors so that the constructs of internalizing and externalizing are equivalent (as indicated by tests of measurement invariance) across gender and religioethnic groups. The identification of these equivalent constructs enables us to compare child internalizing and externalizing behavior problems among subgroups in Mauritius, providing further evidence of the relationships of these constructs to one another, as well as to psychosocial adversity, within a non-Western cultural setting.

Measurement Invariance

Advancements in methodology and statistical software programs have made it possible to statistically test for the equivalence of constructs across groups. Tests of measurement invariance can be conducted at various levels of stringency (see Chen, 2008; Horn & McArdle, 1992; Widaman & Reise, 1997). The least stringent level is configural invariance, which tests for the equivalence of factor structures (i.e., groups have the same number of factors and the same items represent these factors). Lack of configural invariance could occur if a construct is more differentiated or complex in one group than another. The next level is metric invariance, which requires equivalence of factor loadings (i.e., the same units of scale across groups) and the same item structure. In a cross-cultural context, lack of metric invariance could occur if definitions of concepts do not fully overlap across groups, if translations are not exact, or if groups do not use response sets similarly (e.g., one group avoids extreme response options). The third level of invariance is scalar invariance, which tests the equivalence of the intercept (i.e., the same origin of scale across groups). Lack of scalar equivalence could occur if there are consistently higher or lower scores in the groups (e.g., one group provides more desirable or undesirable responses, or the groups use different frames of reference when making ratings). If a measure meets all three forms of invariance, then the scores obtained from this measure (e.g., means, correlations, and variances) can be validly compared across groups.

Tests of measurement invariance are typically conducted by confirmatory factor analysis (CFA). Using CFA, a factor structure that is found previously in one group can be assessed for fit to data from another group. Alternatively, multiple groups can be tested simultaneously to compare model fit across groups; in such multigroup tests, modifications can be made to try to create an invariant structure across groups if one is not initially indicated. In the current study, we used both approaches, first testing a factor structure based on the U.S.-derived CBCL and then modifying it to obtain an invariant factor structure across groups within our sample.

Development of the CBCL

The CBCL was originally designed by Achenbach (1978; Achenbach & Edelbrock, 1979) to measure behavioral problems organized into internalizing, externalizing, and mixed syndromes. Initially, different versions of the scales using different items were derived for boys and girls in three age groups (4-5, 6-11, and 12-16 years) based on separate principal components analyses (Achenbach & Edelbrock, 1983). In 1991, the structure was revised to

reflect the importance of having similar constructs across gender and age groups for assessment, research, and theory. Eight behavior problem syndrome scales were constructed for use with boys and girls from 4 to 18 years old (Achenbach, 1991). A subsequent version of the school-aged CBCL for 6- to 18-year-olds (Achenbach & Rescorla, 2001) replaced six items that had low endorsement or poor psychometric properties with six new items and modified the content of one other item. The eight primary scales of this current version include 103 items, 96 of which appeared on prior versions. Two second-order factors represent Internalizing (comprising Anxious/Depressed, Withdrawn/Depressed, and Somatic Complaints first-order scales) and Externalizing (comprising Rule-Breaking Behavior and Aggressive Behavior first-order scales). The remaining three first-order scales (Attention, Social Problems, and Thought Problems) were not assigned to either Internalizing or Externalizing.

The revisions of the CBCL items and scales over the years were conducted without rigorous tests of strict invariance across groups. This raises the question of how well the current syndrome scales of the CBCL capture similar constructs for boys and girls. In addition, the increasing use of the CBCL in other countries makes it important to ensure these measures derived on U.S. samples assess similar constructs in other cultural settings. This study represents the first study to do this.

Measurement Invariance of the CBCL

Several studies have used the less strict configural invariance to examine the between- and within-culture applicability of the CBCL scales. In a study of more than 58,000 6- to 18-year-olds in 30 societies (Ivanova et al., 2007), configural invariance of the 96-item eight-syndrome structure was largely supported through CFA. All items loaded significantly on their predicted factors in 24 societies, and the hypothesized model had good fit in 26 societies. However, the only African country in this study, Ethiopia, had one of the poorest model fits and had the lowest mean factor loadings of all societies.

Using the same samples plus data from the United States, these authors examined additional indicators of multicultural robustness of the CBCL structure in 6- to 16-year-olds (Rescorla et al., 2007). Internal scale consistencies (i.e., alpha coefficients) were good for the three higher-order factors (Internalizing, Externalizing, and Total Problems) but were only poor to acceptable for six of the seven first-order factors. Across societies, mean scale scores showed greater within-society variation than between-society variation. Within societies, mean level gender differences were found on some scales. For example, girls scored higher than boys on Somatic Complaints in 11 societies and Anxious/Depressed in 9 societies, whereas boys scored higher than girls on Attention Problems in 17 societies, Rule-Breaking Behavior in 14 societies, Aggressive Behavior in 11 societies, and Thought Problems in 2 societies, although effect sizes were small. When examined across age groups, higher internalizing scores were found only in older (age 12-16 years) girls versus boys, and higher externalizing scores were found more consistently for younger (age 6-11 years) boys versus girls than for older boys versus girls. The results of this study suggest relatively consistent trends across cultural settings for both gender and age groups. However, because of the absence of formal tests of measurement invariance, it remains uncertain whether the observed gender differences reflect true differences in these constructs or differences in measurement across and within these cultural settings.

Other studies of the CBCL have used exploratory factor analyses (EFA) to create factor structures that fit data obtained from different cultures. In an early study, De Groot, Koot, and Verhulst (1994) found that a 74-item factor structure fit a Dutch sample; however, half of the items loaded onto different factors from those of the U.S. model (Achenbach, 1991).

Heubeck (2000) compared these two Dutch and U.S. factor structures with one found in an Australian sample. Somatic Complaints, Anxious/Depressed, and Aggressive syndromes exhibited the best convergence across the Dutch, Australian, and American samples, whereas Attention Problems and Social Problems syndromes exhibited low cross-cultural generalizability. Similarly, Berg, Fombonne, McGuire, and Verhulst (1997) found a viable first-order model using 43 items in Dutch and French samples, although this model did not yield a Thought Problems factor and the Delinquent Behavior factor was weak. Weisz et al. (2003) also used EFAs to compare models for Thai and American children. Somatic and Withdrawn factors were the most robust, but the other first-order factors had many items that loaded on different factors across Thai and American boys and girls. Overall, these studies indicate that the first-order syndromes can generally be identified across cultures but specific items and factor loadings vary across culture. In addition, some factors (e.g., Somatic Complaints) appear more robust across cultures than others (e.g., Delinquency).

To our knowledge, no prior study has examined more stringent tests of measurement invariance, including metric and scalar invariance, of the CBCL factor structure across societies or within a society by gender, ethnicity, or other subgroups. A study by Guttmannova, Szanyi, and Cali (2007) investigated measurement invariance for a subset of items from the Behavior Problems Index (Peterson & Zill, 1986), a measure comprising 28 items that are mostly identical to CBCL items. Using 17 of these items to form Internalizing and Externalizing factors, they found measurement invariance of factor loadings and thresholds across African American, Caucasian American, and Hispanic American children from the National Longitudinal Survey of Youth, providing support for the validity of these two factors for cross-group comparisons.

Current Study

The current study assessed the fit of the internalizing and externalizing syndrome structure of the CBCL in a sample from Mauritius, an island country with much cultural and ethnic diversity. Mauritius is located approximately 500 miles east of Madagascar in the Indian Ocean. It was first settled by the Dutch at the end of the 16th century, then was a French colony from 1715 to 1810, and finally a British colony until it gained its independence in 1968. The majority of the current population are descendents of slaves brought from Africa (about 25%, Creole) and indentured servants and traders from India and Pakistan (about 71%). Indian Mauritians further identify themselves by their religion as Hindu (about 52%) and Muslim (about 19%). Thus, the Mauritian culture includes European, Indian, and African influences that may be associated with varying perceptions and/or manifestations of child behavior problems.

CBCL data drawn from this society give us the opportunity to explore the generalizability of the CBCL constructs to a non-Western setting. In addition, differences across the three primary religioethnic groups, Creoles, Hindus, and Muslims, allow us to test for differences in the CBCL factor structure among subgroups within this society. Thus, the Mauritius data allow us to identify items within the CBCL that do and do not operate well in a cultural context that is distinct from that of the United States and to examine the degree to which there is support for invariance of these constructs by gender and religioethnicity within this non-Western culture.

Method

Participants

Data are from the Joint Child Health Project, an ongoing longitudinal study in Mauritius (see Raine, Liu, Venables, Mednick, & Dalais, 2010, for details). All families with children born

during a 1-year period in 1969 to 1970 in two towns on the island, Quatre Bornes and Vacoas, were recruited into the study, and 100% ($N = 1,795$) agreed to participate. These towns were chosen in part because of their representativeness of the ethnic distribution of the entire country.

When the children were 11 years old ($M = 11.0$ years, $SD = 0.66$), 1,213 parents completed the CBCL in Kreol, the common spoken language of Mauritius. The English version of the CBCL was translated by native Mauritians into Kreol and then back-translated into English until semantic equivalence of the instruments was obtained. Bilingual Mauritian research staff verbally administered the Kreol CBCL to parents and clarified item content as appropriate. The sample for the current analysis was derived from the 1,146 of 1,213 children interviewed who were from the three largest religioethnic groups: Creole (25%), Hindu (53%), and Muslim (22%); 45% to 47% of each religioethnic group was female.

A measure of psychosocial adversity was calculated for each child by adding one point for each of the following 14 variables rated by a social worker during a home visit: living in poor housing, rented accommodation, overcrowded home, no electricity or water, no television, child has neither good toys nor books, father uneducated, mother uneducated, parent psychiatrically ill, parent physically ill, teenage mother, single parent status, separation from both parents, and five or more siblings (see Liu, Raine, Venables, & Mednick, 2004; Raine, Reynolds, Venables, Mednick, & Farrington, 1998). The three religioethnic groups did not differ on mean levels of psychosocial adversity at age 11, $F(2, 1,099) = 0.95$, $p = .39$.

CBCL Measure

We used the original school-aged version of the CBCL and mapped these items onto six of the eight current U.S.-derived behavioral syndrome scales that reflect higher-order constructs of Internalizing (Anxious/Depressed, Withdrawn/Depressed, Somatic Complaints) and Externalizing (Rule-Breaking Behavior, Aggressive Behavior), as well as Attention Problems, which is thought to be a mixture of Internalizing and Externalizing. We omitted Social Problems and Thought Problems from our analyses because we wanted to focus on the syndromes more consistently identified with the internalizing and externalizing constructs.

Analyses were based on the 57 items included in these six syndrome scales for both genders in the original version of the CBCL. This excluded 14 items administered only to boys or to girls and the 6 new items and 1 revised item in the 2001 CBCL version. In addition, because of low frequency of endorsement of 2 items with similar content (“steals at home” and “steals outside of the home”), we parceled these 2 items into a single item (“steals”). The 57 items for the six primary factors are shown in Table 1. The original three-level categorical response options (not true, somewhat or sometimes true, very true or often true) were recoded into dichotomous categories (no endorsement vs. endorsement) because of low use of the highest category (e.g., 17 items were endorsed *very true or often true* by fewer than 5% of the sample). Dichotomous scoring is consistent with recent studies of the factor structure of the CBCL because of consistently low endorsement of the upper category (see Ivanova et al., 2007).

Analyses

Analyses were conducted using Mplus 6.1 (Muthén & Muthén, 2010). We constructed the primary factor structure by loading items onto six correlated factors based on the U.S.-derived syndrome factors (see Table 1). To meet criteria for measurement invariance, the model had to show evidence of configural invariance of the factor structure and metric/

scalar invariance across both gender (boys, girls) and religioethnic (Creole, Hindu, Muslim) groupings.

We utilized the recommended weighted least square means estimation for testing categorical variables, with factor analyses based on matrices of tetrachoric correlations (Muthén & Muthén, 2010). We examined the fit of the initial model using the root mean square error of approximation (RMSEA; Steiger, 1990) and comparative fit index (CFI; Bentler, 1990). We considered an RMSEA of .08 and lower and a CFI of .90 and higher to represent adequate fit (Browne & Cudeck, 1993). The DIFFTEST procedure was then used to produce chi-square difference tests that compared the fits of nested multigroup models that had item loadings and thresholds constrained (metric/scalar invariance) versus freed (configural invariance) across groups. When the χ^2 of the DIFFTEST was nonsignificant (i.e., $p > .05$), this indicated support for measurement invariance (i.e., the more restrictive metric/scalar invariant model did not fit significantly worse than the less restrictive configural invariant model). Note that when using dichotomous indicators, metric and scalar invariance are tested in a single step because of the non-independence of the mean and variance (see Muthén & Muthén, 2010).

When the factor structure did not show support for measurement invariance, we proceeded to use a stepwise process to establish measurement invariance of the CBCL primary factor behavior syndromes and then the secondary internalizing and externalizing factors. Items that did not show configural invariance (i.e., that had either zero or negative loadings on their purported factors) were first removed. We then proceeded to test invariance at the metric/scalar level. We first examined each primary factor separately and removed items iteratively in order of highest modification index (MI, which indicates the χ^2 value a parameter contributes to the model misfit) until a nonsignificant change in model χ^2 indicated support for measurement invariance. These refined individual factors were then placed back in the multifactor model, allowing the primary factors to be correlated in a multifactor model and to test the fit of each factor when estimated in conjunction with other indicators of internalizing and externalizing problems. We employed unit variance identification (UVI; factor variance fixed at 1 in this stage) so that every item loading could contribute to the misfit of the model. Items were removed until support for measurement invariance by gender and by religioethnicity was achieved. These two models were then combined to find support for measurement invariance for both grouping variables in a single model. Last, items retained in the final primary factor model were placed in a confirmatory model of Internalizing and Externalizing secondary factors, and we tested for measurement invariance of this structure for both grouping variables.

Following the establishment of measurement invariance of both the primary and secondary factor models, we used these scales to compare gender and religioethnic groups on factor scores (using the p value generated by Mplus) and correlations (using the Fisher r -to- z transformation). We also examined psychosocial adversity as a potential mediator of group differences on behavior problem factor scores within each religioethnic group. We ran two sets of multivariate analyses of covariance (MANCOVAs) for the primary scales and the secondary scales that included psychosocial adversity, gender, and religioethnicity as predictors of the CBCL scale scores.

Results

Table 1 shows means and standard deviations by gender and religioethnicity for the 57 items used in the CFA models. For items that differed across gender, girls tended to have higher scores on the internalizing items and boys tended to have higher scores on the externalizing items ($p < .01$ to adjust for multiple comparisons). For items that differed by

religioethnicity, Creoles tended to have higher scores than Hindus and Muslims; the exception was for items on the Aggressive Behavior scale, where the pattern of differences was more variable.

Configural Invariance

Table 2 displays the factor loadings for the initial configural invariance model using all 57 items. Because we found a high correlation between Rule-Breaking and Aggressive Behavior in Creoles ($r = .99$) and Hindus ($r = .98$) in initial models, these two factors were combined into one in the religioethnicity model. Three items, “must be perfect,” “feels guilty,” and “loud,” had zero or significant negative loadings on their factors and were removed. Models based on the remaining 54 items had adequate fit as indicated by the RMSEA for both gender (RMSEA = .03, CFI = .87) and religioethnicity (RMSEA = .03, CFI = .81), although they were slightly below acceptable for the CFI. These item loading and fit values indicated the U.S.-derived factor structure was generally supported for the Mauritian data at the configural level of invariance.

Metric/Scalar Invariance

The metric/scalar invariance models of the 54-item factor structure had poorer fit than the configural invariance models for gender, $\chi^2(42) = 173.9, p < .001$, and for religioethnicity, $\chi^2(88) = 146.2, p < .001$, indicating the factors do not measure the same constructs across groups. Thus, we next sought to obtain measurement invariance of the primary and secondary factors as indicated by metric/scalar invariance (referred to as simply *invariance* in this section).

Refinement of the Individual Primary Factors—In this step, each factor was examined separately to determine which items contributed to a lack of invariance. Across gender, Somatic Complaints, $\chi^2(6) = 9.2, p = .16$, and Rule-Breaking, $\chi^2(3) = 3.7, p = .30$, had initial support for invariance. Anxious/Depressed was invariant after removing “cries,” $\chi^2(7) = 13.6, p = .06$, Attention Problems after removing “day-dreams,” $\chi^2(5) = 2.4, p = .79$, Withdrawn/Depressed after removing “shy” and “secretive,” $\chi^2(3) = 3.3, p = .35$, and Aggressive Behavior after removing “demands attention,” “cruel to others,” “sulks,” and “tantrums,” $\chi^2(10) = 17.2, p = .07$.

Across religioethnic groups, Anxious/Depressed, $\chi^2(14) = 22.4, p = .07$, Attention Problems, $\chi^2(12) = 15.5, p = .22$, and Somatic Complaints, $\chi^2(12) = 20.6, p = .06$, had initial support for invariance. Withdrawn/Depressed was invariant after removing “secretive,” $\chi^2(8) = 12.3, p = .14$, and Rule-Breaking/Aggressive after removing “argues,” “bad friends,” “disobedient at school,” “swears,” and “fights,” $\chi^2(28) = 40.8, p = .06$.

Reconstruction of the Primary Factor Models—We then reconstructed the multifactor primary models using the refined primary factors obtained in the previous step. Initial runs of the reconstructed gender model had the same high correlation between the Rule-Breaking and Aggressive Behavior factors in females ($r = .99$) that was previously seen among Creoles and Hindus, so these two factors were collapsed into a Rule-Breaking/Aggressive factor. In this five-factor model, four items had substantial cross-loadings on additional factors: “hyperactive,” “moody,” “unhappy,” and “worries.” Removing these items resulted in an invariant factor structure, $\chi^2(69) = 79.8, p = .18$. For religioethnicity, four items were removed to achieve an invariant structure: “lies/cheats,” “nausea,” “steals,” and “suicidal talk,” $\chi^2(138) = 164.1, p = .06$.

Final Primary Factor Model—The final step for establishing a primary factor invariant model was to create a model that was invariant for both gender and religioethnicity. This

required excluding 21 of the original 57 items, including 12 removed in the gender model, 9 removed in the religioethnicity model, and 1 removed in both. However, after eliminating these items, only one Rule-Breaking item, “runs away,” remained on the Rule-Breaking/Aggressive Behavior factor. Because including this item changed the nature of a factor that otherwise measured solely aggressive behavior, we dropped it in subsequent analyses and once again we referred to this factor as Aggressive Behavior.

Without any further modifications, this model was invariant for both gender, $\chi^2(59) = 69.6$, $p = .16$, and religioethnicity, $\chi^2(118) = 138.6$, $p = .09$. Table 3 displays the parameter estimates of the final primary factor model freed (configural) and fixed (metric/scalar) for gender and for religioethnicity. Reliabilities for the latent constructs (H) were between .70 and .98, indicating good reliability (Hancock & Mueller, 2001).

Secondary Factor Models—We evaluated the hypothesized Internalizing and Externalizing factor structure by running CFAs with the items retained in the final primary factor model. Internalizing was composed of items from the Anxious/Depressed, Withdrawn/Depressed, and Somatic Complaints factors and Externalizing was composed of the Aggressive Behavior items. Given the complex nature of the six Attention Problems items, these were initially allowed to load on both factors; results indicated “confused,” “can’t concentrate,” and “stares blankly” loaded more highly onto Internalizing, and “impulsive” and “poor schoolwork” loaded more highly onto Externalizing. These items were then loaded onto these factors; “acts too young” loaded onto neither factor, so it was not retained in the secondary model.

This Internalizing-Externalizing model was invariant without modification for both gender, $\chi^2(60) = 75.4$, $p = .09$, and religioethnicity, $\chi^2(120) = 140.5$, $p = .10$. Table 4 displays the estimates of this final secondary factor model by gender and religioethnic group. Reliabilities for these latent constructs were good, ranging from .85 to .91.

Gender and Religioethnic Group Differences

Having established measurement invariance of the refined factor structures, we could now compare the scores of the gender and religioethnicity groups. On these pared-down primary factors, girls were higher than boys on Somatic Complaints (standardized latent mean difference of 0.21, $p < .01$) and lower on Aggressive Behavior (-0.32 , $p < .001$) and Attention Problems (-0.20 , $p < .05$; see Table 3). Creole children were higher than Hindu and Muslim children on all five factors (see Table 3 for latent mean differences, ranging from 0.35 to 0.87, all $ps < .001$). Hindu children were higher than Muslims on Somatic Complaints (-0.65 vs. -0.87 , $p < .05$) and lower on Aggressive Behavior (-0.76 vs. -0.53 , $p < .05$). The correlations among the factors ranged from .35 to .93 and did not significantly differ across gender or religioethnic groups.

On the refined secondary factors, girls were lower than boys on Externalizing (-0.32 , $p < .001$), and Creoles were higher on both Internalizing and Externalizing than Hindus and Muslims (see Table 4 for latent mean differences, ranging from 0.55 to 0.81, $ps < .001$). The correlation between Internalizing and Externalizing ranged from .61 to .79 for all groups and did not significantly differ across gender or religioethnicity. The pattern of gender differences was consistent across the religioethnic groups (detailed results available on request).

Greater psychosocial adversity was associated with higher scores on all five primary and both secondary scales, but religioethnicity differences in behavior problems were not explained by group differences in adversity. The addition of psychosocial adversity and gender in the models reduced the effect size of religioethnicity by only .001 in both the

primary factor (η^2 from .055 to .054) and secondary factor (η^2 from .049 to .048) models. The pattern of relative mean differences among the religioethnic groups also remained consistent with and without psychosocial adversity and gender in the models. Thus, the religioethnic mean differences on the factors were relatively robust and were not accounted for by psychosocial adversity or gender.

Discussion

This study had two goals: to examine the measurement structure of the CBCL in a society very different from the one for which it was designed and to create an equivalent factor structure based on this structure to compare gender and religioethnic group scores within a non-Western society. We found support for configural invariance of the U.S.-derived factor structure for all but three items, which suggests the initial structure of the model was relatively good for this Mauritian sample. Along with prior studies conducted in other societies, this indicates that the first-order syndromes can generally be identified across cultures, with some variation in items and factor loadings. There was not support, however, for this structure meeting the stricter requirements of metric/scalar invariance across groups within our sample. Thus, we refined the standard CBCL syndrome primary factors for internalizing, externalizing, and attention problem behaviors by removing items that contributed to a lack of invariance. The final refined models demonstrated support for measurement invariance at the configural, metric, and scalar levels for both gender and religioethnicity and for both primary and secondary factors.

Our scales are pared-down versions of the U.S.-derived CBCL syndrome scales and there might be concern that they do not represent the same constructs. However, the correlations among the factors are similar to those reported in the CBCL manual in the United States (Achenbach, 1991), and no additional items were included as indicators on these scales. Moreover, many of the items removed in our sample were relatively poor indicators in earlier U.S. factor structures, as indicated by lower factor loadings on their scales (e.g., “shy,” “swears,” “underactive”) and loading on multiple scales (e.g., “cruel to others,” “moody,” “sulks”) in 6 to 11-year-old boys and girls (Achenbach & Edelbrock, 1983). Thus, our invariant factors, albeit reduced, appear to generally capture similar constructs as those identified by the original CBCL factors.

The final measurement invariant model showed support for five of the six primary syndrome factors of the CBCL that we examined. We were unable to identify a Rule Breaking factor, but this seems likely because of the few items we had for this factor (five) rather than being a statement about the unique nature of rule breaking behavior in Mauritius. In fact, our endorsement rates of these items were within 5% of those reported in the CBCL manual for 4- to 11-year-old boys and girls (Achenbach, 1991) for all items except “bad friends,” which was higher in our sample (boys 49% vs. 13%, girls 39% vs. 5%). Other studies conducted with non-American samples also have not found a separate Delinquency scale (Berg et al., 1997; Viola et al., 2011). Age 11 is relatively young for the emergence of delinquent behaviors, and such behaviors at this age may be due more to short-lived environmental pressures (e.g., peers, imitation) rather than to a cohesive behavior pattern, which may emerge later in adolescents as distinct from aggressive behavior (see Dishion & Tipsord, 2011; Loeber & Hay, 1997; Moffitt, 1993; Steinberg et al., 2006). It is also possible the types of behaviors indicative of delinquency are more varied across cultures than are other components of internalizing and externalizing and the items in the CBCL do not capture how delinquency is manifested outside Western societies.

In contrast to the differences seen with Rule Breaking, the Somatic Complaints factor required very little modification to obtain invariance across gender and religioethnic groups

in our sample; only the “nausea” item needed to be dropped to obtain invariance. This finding is consistent with other studies showing strong cross-cultural robustness of this scale (Berg et al., 1997; DeGroot et al., 1994; Weisz et al., 2003), perhaps because somatic symptoms are more often reported by the children to their parents, are easier behaviors for parents to identify, or are more universally experienced by children than the behaviors indexed by other factors.

The Mauritian data also fit the hypothesized secondary factor structure of the CBCL. Invariance of the refined Internalizing and Externalizing factors was found for both gender and religioethnicity without requiring the removal of other items, except one Attention item, “acts too young,” that did not load on either factor. The remaining five Attention items, however, did fit into the secondary factor structure. The Attention scale is proposed to be composed of both internalizing and externalizing behaviors and therefore is generally not included in second-order factor models (Achenbach, 1991). Our findings indicate that Attention as a first-order scale would not load well onto a second-order factor model, but that these items did fit in a model that loads individual items directly onto secondary factors. The Attention items loaded in a consistent pattern that made sense theoretically; items that are more indicative of inattention (“confused,” “can’t concentrate,” “stares blankly”) loaded on Internalizing, whereas “impulsive” and “poor schoolwork” loaded on Externalizing. Similar results may be found in other societies if secondary factors are tested or modified in the same manner.

Gender and Religioethnic Group Differences

In our sample of 11-year-old Mauritian children, girls were rated higher by their parents than boys on Somatic Complaints and lower on Aggressive Behavior, Attention Problems, and Externalizing. These results are generally consistent with gender differences reported for younger age groups (6-11 years old) in other societies (Rescorla et al., 2007). Across societies, younger girls consistently are lower on externalizing problems than boys, and no gender differences are found on internalizing behavior. In older children (12-16 years old), gender differences for externalizing behavior are less consistent, but girls often score higher than boys on internalizing problems. The original CBCL scale construction acknowledged the importance of age 11 for potential biobehavioral developmental shifts and proposed different factor structures for 6- to 11-year-old and 12- to 16-year-old boys and girls (Achenbach, 1978; Achenbach & Edelbrock, 1979). In Mauritius, most 11-years-olds are in their sixth year of schooling and are still in primary school, so they may be more likely to resemble the younger developmental period seen in other societies. However, age 11 is the upper limit of this age range so it is not surprising that some gender differences in internalizing behaviors (e.g., Somatic Complaints) do emerge in this sample that are more consistent with older children in other societies.

Religioethnic group differences also were found in our sample. Creole children were rated higher by their parents on all five primary factors and both secondary factors compared with children in the other two groups. These findings were not accounted for by psychosocial adversity, although greater psychosocial adversity did uniquely predict higher levels of all behavior problem syndromes. These findings were also not accounted for by gender differences within groups, given that patterns of gender differences were consistent across religioethnic groups.

The reasons for these religioethnic group differences are not clear. Parent reports of problem behaviors in their children are likely influenced by a number of factors, including their perceptions, observation of their children relative to other children, cultural stigmas attached to different behavior problems, and their manner of responding to questionnaires. We do not have CBCL ratings from other informants to cross validate differences, which would help

address if these group differences are attributable to rater bias, true differences, or a combination of the two. However, when compared with U.S. 4- to 11-year-old boys and girls on item-level endorsement rates (Achenbach, 1991), some patterns do emerge. Mauritian parents were more likely to endorse internalizing behaviors in their children than U.S. parents, with a few exceptions (e.g., “self-conscious,” “rather be alone,” “stomachaches”). On aggressive behaviors, however, Creole endorsement rates were generally similar to U.S. rates, whereas Hindu and Muslim rates were lower; on a few items, however, Creoles were higher than U.S. rates (e.g., “destroys own things,” “teases”). Attention item endorsement rates were somewhat mixed but generally were higher in all three Mauritian groups than in the U.S. sample.

Lower endorsement rates by Hindu and Muslim parents relative to Creole parents (and to U.S. parents for aggressive behaviors) may be related to the Asian idea of “saving face” (Yabuuchi, 2004), whereby Indian parents would be less likely to endorse problems in their children than parents from other groups because they are worried that acknowledging their children’s problem behaviors might negatively affect their family’s reputation. Given Creole children were rated by their parents as having more problems on all factors, this suggests social desirability bias may be one source for these differences.

Among Indian children, latent factor means were generally similar for Hindu and Muslim children. Significant differences were found on only two primary factors, with Hindus having higher scores on Somatic Complaints and lower scores on Aggressive Behavior than Muslims. The reasons for these differences also are not clear, but influences of parenting practices, parental expectations, and group norms are avenues to pursue in future research.

Limitations and Conclusions

There are several limitations associated with the study design that should be considered when interpreting our results. First, our sample was limited to 11-year olds, and therefore we did not examine age-related trends. This narrow age range may have contributed to our lack of finding a Rule Breaking factor, which could have emerged in a sample with a broader age range. It also, however, provided an opportunity to focus on gender and religioethnic groups and largely removed the expected variation in factor structure and item endorsement associated with developmental differences across ages. A second limitation concerns the accuracy of the CBCL translation, as is the case in all research applying measures created in one society to another society. Although we attempted to minimize differences in meaning using a back-translation procedure with bilingual individuals, subtle language differences may have introduced some variations. Nonetheless, all participants in this study were administered the same translated measure, so any unreliability is likely to be random rather than systematic across groups. Third, the use of an older version of the CBCL that consisted only of items included on both the original boy and girl structures reduced our initial item set for these five scales from 77 to 57 items. A full set of items could have increased our ability to identify the rule breaking factor and would have provided a better comparison with studies conducted in other societies. Our removal of items also may have altered the meaning of the constructs from those of the original CBCL scales. Finally, it is possible that the CBCL items we studied did not fully capture internalizing and externalizing problem behavior constructs in the Mauritian setting (i.e., incomplete item coverage). Additional items may be required to capture the full nature of these behavioral syndromes in this non-Western context.

Despite these limitations, this study found invariant internalizing and externalizing primary and secondary factors using items from the CBCL behavior syndrome scales in a large cohort sample from the African country of Mauritius. There are more than 7,000 publications using the CBCL, but we know of no other study that has examined its factor

structure at the levels of metric and scalar invariance. This study represents the first known study to demonstrate strict measurement invariance of internalizing and externalizing factors using the CBCL. These refined factors showed a pattern of gender differences in a non-Western culture that were consistent with variations found in studies of children of this age group in other cultural settings. In addition, our finding of religioethnic group differences in mean factor scores suggests a further area to explore in understanding how internalizing and externalizing behaviors are manifested and perceived within these cultural subgroups.

Acknowledgments

We thank the staff of the Joint Child Health Project for their assistance with data collection and management.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by grants from the National Institutes of Health (K08 AA14265 and R01 AA18179), Mauritian Ministry of Health, UK Medical Research Council, and Wellcome Trust.

References

- Achenbach TM. The Child Behavior Profile: I. Boys aged 6-11. *Journal of Consulting and Clinical Psychology*. 1978; 46:478–488. doi:10.1037/0022-006X.46.3.478. [PubMed: 670491]
- Achenbach, TM. *Manual for the Child Behavior Checklist: 4-18 and 1991 Profile*. University of Vermont, Department of Psychiatry; Burlington: 1991.
- Achenbach TM, Edelbrock CS. The Child Behavior Profile: II. Boys aged 12-16 and girls aged 6-11 and 12-16. *Journal of Consulting and Clinical Psychology*. 1979; 47:223–233. doi: 10.1037/0022-006X.47.2.223. [PubMed: 469068]
- Achenbach, TM.; Edelbrock, C. *Manual for the Child Behavior Checklist and Revised Child Behavior Profile*. University of Vermont, Department of Psychiatry; Burlington: 1983.
- Achenbach, TM.; Rescorla, LA. *Manual for the ASEBA School-Age Forms & Profiles*. University of Vermont, Research Center for Children, Youth, and Families; Burlington: 2001.
- Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin*. 1990; 107:238–246. [PubMed: 2320703]
- Berg I, Fombonne E, McGuire R, Verhulst F. A cross-cultural comparison of French and Dutch disturbed children using the Child Behavior Checklist (CBCL). *European Child and Adolescent Psychiatry*. 1997; 6:7–11. [PubMed: 9112041]
- Berry JW. Imposed etics, etics, derived etics. *International Journal of Psychology*. 1989; 24:721–735.
- Browne, MW.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA.; Long, JS., editors. *Testing structural equation models*. Sage; Newbury Park, CA: 1993. p. 136-162.
- Chen FF. What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*. 2008; 95:1005–1018. doi:10.1037/a0013193. [PubMed: 18954190]
- De Groot A, Koot HM, Verhulst FC. Cross-cultural generalizability of the Child Behavior Checklist cross-informant syndromes. *Psychological Assessment*. 1994; 6:225–230.
- Dishion TJ, Tipsord JM. Peer contagion in child and adolescent social and emotional development. *Annual Review of Psychology*. 2011; 62:189–214.
- Eaton NR, Keyes KM, Krueger RF, Balsis S, Skodol AE, Markon KE, Hasin DS. An invariant dimensional liability model of gender differences in mental disorder prevalence: Evidence from a national sample. *Journal of Abnormal Psychology*. 2012; 121:282–288. doi:10.1037/a0024780. [PubMed: 21842958]
- Guttmanova K, Szanyi JM, Cali PW. Internalizing and externalizing behavior problem scores: Cross-ethnic and longitudinal measurement invariance of the Behavior Problem Index. *Educational and Psychological Measurement*. 2008; 68:676–694. doi:10.1177/0013164407310127.
- Hancock, GR.; Mueller, RO. Rethinking construct reliability within latent variable systems. In: Cudeck, R.; du Toit, S.; Sörbom, D., editors. *Structural equation modeling: Present and future: A*

- festschrift in honor of Karl Jöreskog. Scientific Software International; Lincolnwood, IL: 2001. p. 195-216.
- Horn JL, McArdle JJ. A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*. 1992; 18:117–144. [PubMed: 1459160]
- Heubeck BG. Cross-cultural generalizability of CBCL syndromes across three continents: From the USA and Holland to Australia. *Journal of Abnormal Child Psychology*. 2000; 28:439–450. [PubMed: 11100918]
- Ivanova MY, Achenbach TM, Dumenci L, Rescorla LA, Almqvist F, Weintraub S, Verhulst FC. Testing the 8-syndrome structure of the Child Behavior Checklist in 30 societies. *Journal of Clinical Child and Adolescent Psychology*. 2007; 36:405–417. doi:10.1080/10705519909540118. [PubMed: 17658984]
- Liu J, Raine A, Venables PH, Mednick SA. Malnutrition at age 3 years and externalizing behavior problems at ages 8, 11, and 17 years. *American Journal of Psychiatry*. 2004; 161:2005–2013. doi:10.1176/appi.ajp.161.11.2005. [PubMed: 15514400]
- Loeber R, Hay D. Key issues in the development of aggression and violence from childhood to early adulthood. *Annual Review of Psychology*. 1997; 48:371–410.
- McLaughlin KA, Hilt LM, Nolen-Hoeksema S. Racial/ethnic differences in internalizing and externalizing symptoms in adolescents. *Journal of Abnormal Child Psychology*. 2007; 35:801–816. [PubMed: 17508278]
- Moffitt TE. Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review*. 1993; 100:674–701. [PubMed: 8255953]
- Muthén, LK.; Muthén, BO. *Mplus user's guide*. 6th ed.. Muthén & Muthén; Los Angeles, CA: 2010.
- Peterson JL, Zill N. Marital disruption, parent-child relationships, and behavior problems in children. *Journal of Marriage and the Family*. 1986; 48:295–307.
- Raine A, Reynolds C, Venables PH, Mednick SA, Farrington DP. Fearlessness, stimulation-seeking, and large body size at age 3 years as early predispositions to childhood aggression at age 11 years. *Archives of General Psychiatry*. 1998; 55:745–751. [PubMed: 9707386]
- Raine A, Liu J, Venables PH, Mednick SA, Dalais C. Cohort profile: The Mauritius Child Health Project. *International Journal of Epidemiology*. 2010; 39:1441–1451. doi:10.1093/ije/dyp341. [PubMed: 19995862]
- Rescorla L, Achenbach T, Ivanova MY, Dumenci L, Almqvist F, Bilenberg N, Verhulst F. Behavioral and emotional problems reported by parents of children ages 6 to 16 in 31 societies. *Journal of Emotional and Behavioral Disorders*. 2007; 15:130–142. doi:10.1037/1040-3590.6.4.304.
- Steiger JH. Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*. 1990; 25:173–180.
- Steinberg, L.; Dahl, R.; Keating, D.; Kupfer, DJ.; Masten, AS.; Pine, DS. The study of developmental psychopathology in adolescence: integrating affective neuroscience with the study of context. In: Cicchetti, D.; Cohen, DJ., editors. *Developmental psychopathology: Developmental neuroscience*. 2nd ed.. Vol. 2. Wiley; Hoboken, NJ: 2006. p. 710-741.
- Viola L, Garrido G, Rescorla L. Testing multicultural robustness of the Child Behavior Checklist in a national epidemiological sample in Uruguay. *Journal of Abnormal Child Psychology*. 2011; 39:897–908. doi:10.1007/s10802-011-9500-z. [PubMed: 21424940]
- Weisz JR, Weiss B, Suwanlert S, Chaiyasit W. Syndromal structure of psychopathology in children of Thailand and the United States. *Journal of Consulting and Clinical Psychology*. 2003; 71:375–385. doi:10.1037/0022-006x.71.2.375. [PubMed: 12699031]
- Widaman, KF.; Reise, SP. Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In: Bryant, KJ.; Windle, M.; West, SG., editors. *The science of prevention: Methodological advances from alcohol and substance abuse research*. American Psychological Association; Washington, DC: 1997. p. 281-324.
- Yabuuchi A. Face in Chinese, Japanese, and U.S. American cultures. *Journal of Asian Pacific Communication*. 2004; 14:261–297.

Table 1

Means and Standard Deviations for Dichotomized Child Behavior Checklist Syndrome Items by Gender and Religioethnicity.

	Gender		Religioethnicity		
	Boys (<i>n</i> = 623), <i>M</i> (<i>SD</i>)	Girls (<i>n</i> = 523), <i>M</i> (<i>SD</i>)	Creole (<i>n</i> = 286), <i>M</i> (<i>SD</i>)	Hindu (<i>n</i> = 605), <i>M</i> (<i>SD</i>)	Muslim (<i>n</i> = 255), <i>M</i> (<i>SD</i>)
Anxious/Depressed					
Cries	.20 _a (.40)	.29 _b (.45)	.33 _a (.47)	.21 _b (.41)	.21 _b (.41)
Fears school	.06 (.24)	.05 (.21)	.08 (.27)	.04 (.20)	.06 (.23)
Fears doing bad	.39 (.49)	.41 (.49)	.39 (.49)	.42 (.49)	.38 (.49)
Must be perfect	.82 (.38)	.83 (.37)	.85 (.36)	.84 (.37)	.77 (.42)
Feels unloved	.15 _a (.36)	.22 _b (.42)	.32 _a (.47)	.15 _b (.35)	.12 _b (.32)
Feels worthless	.18 (.39)	.18 (.39)	.25 _a (.43)	.16 _b (.36)	.17 _{ab} (.38)
Nervous	.55 (.50)	.50 (.50)	.66 _a (.47)	.48 _b (.50)	.47 _b (.50)
Anxious	.27 _a (.44)	.35 _b (.48)	.41 _a (.49)	.28 _b (.45)	.24 _b (.43)
Feels guilty	.67 (.47)	.69 (.46)	.67 (.47)	.69 (.46)	.67 (.47)
Self-conscious	.26 (.44)	.22 (.42)	.32 _a (.47)	.23 _b (.42)	.20 _b (.40)
Talks about suicide	.04 (.19)	.03 (.17)	.05 (.22)	.03 (.16)	.03 (.18)
Worries	.19 (.40)	.21 (.41)	.28 _a (.45)	.17 _b (.37)	.19 _{ab} (.40)
Withdrawn/Depressed					
Rather be alone	.22 (.42)	.19 (.39)	.28 _a (.45)	.18 _b (.39)	.18 _{ab} (.39)
Won't talk	.07 (.25)	.10 (.30)	.13 (.33)	.07 (.25)	.07 (.26)
Secretive	.29 (.45)	.24 (.43)	.40 _a (.49)	.23 _b (.42)	.21 _b (.41)
Shy	.10 _a (.30)	.47 _b (.50)	.28 (.45)	.27 (.45)	.24 (.43)
Underactive	.37 (.48)	.44 (.50)	.43 (.50)	.39 (.49)	.39 (.49)
Unhappy	.25 _a (.43)	.33 _b (.47)	.33 (.47)	.29 (.45)	.23 (.42)
Withdrawn	.16 (.37)	.18 (.38)	.22 (.41)	.15 (.35)	.16 (.37)
Somatic Complaints					
Nightmares	.15 (.36)	.18 (.39)	.25 _a (.44)	.15 _b (.36)	.10 _b (.30)
Dizzy	.28 (.45)	.28 (.45)	.41 _a (.49)	.25 _b (.43)	.21 _b (.41)
Overtired	.41 (.49)	.43 (.50)	.57 _a (.50)	.37 _b (.48)	.35 _b (.48)
Aches	.25 (.43)	.31 (.46)	.40 _a (.49)	.25 _b (.43)	.22 _b (.42)
Headaches	.25 _a (.44)	.32 _b (.47)	.36 _a (.48)	.28 _{ab} (.45)	.20 _b (.40)
Nausea	.03 (.18)	.05 (.22)	.07 (.26)	.03 (.17)	.03 (.18)
Stomachaches	.04 (.20)	.07 (.26)	.08 (.27)	.05 (.21)	.05 (.21)
Vomits	.04 _a (.20)	.08 _b (.28)	.08 (.28)	.05 (.23)	.05 (.22)
Rule-Breaking Behavior					
Bad friends	.49 _a (.50)	.39 _b (.49)	.44 (.50)	.44 (.50)	.47 (.50)
Lies/cheats	.28 (.45)	.22 (.42)	.47 _a (.50)	.19 _b (.39)	.17 _b (.38)
Runs away	.05 (.21)	.03 (.17)	.07 (.26)	.02 (.16)	.03 (.18)
Steals	.02 (.15)	.01 (.11)	.04 (.20)	.01 (.09)	.02 (.13)

	Gender			Religioethnicity	
	Boys (<i>n</i> = 623), <i>M</i> (<i>SD</i>)	Girls (<i>n</i> = 523), <i>M</i> (<i>SD</i>)	Creole (<i>n</i> = 286), <i>M</i> (<i>SD</i>)	Hindu (<i>n</i> = 605), <i>M</i> (<i>SD</i>)	Muslim (<i>n</i> = 255), <i>M</i> (<i>SD</i>)
Swears	.19 _a (.40)	.12 _b (.32)	.21 (.41)	.15 (.35)	.14 (.35)
Aggressive Behavior					
Argues	.71 (.45)	.31 (.46)	.81 _a (.39)	.62 _b (.48)	.73 _a (.45)
Cruel to others	.47 _a (.50)	.41 _b (.49)	.47 _a (.50)	.36 _b (.48)	.37 _{ab} (.48)
Demands attention	.46 _a (.50)	.25 _b (.43)	.52 (.50)	.46 (.50)	.54 (.50)
Destroys own things	.25 _a (.43)	.18 _b (.38)	.33 _a (.47)	.18 _b (.38)	.20 _a (.40)
Destroys others' things	.08 (.27)	.07 (.25)	.14 _a (.34)	.04 _b (.19)	.09 _{ab} (.29)
Disobedient at home	.31 _a (.46)	.04 _b (.18)	.42 _a (.50)	.22 _b (.41)	.25 _b (.44)
Disobedient at school	.05 (.22)	.04 (.18)	.07 _a (.26)	.04 _{ab} (.20)	.02 _b (.13)
Fights	.27 (.44)	.23 (.42)	.30 (.46)	.22 (.41)	.26 (.44)
Hits people	.09 (.29)	.07 (.26)	.12 (.32)	.06 (.24)	.09 (.29)
Screams	.56 (.50)	.50 (.50)	.63 _a (.48)	.48 _b (.50)	.53 _{ab} (.50)
Stubborn	.48 (.50)	.42 (.49)	.61 _a (.49)	.40 _b (.49)	.41 _b (.49)
Moody	.30 (.46)	.31 (.46)	.36 (.48)	.28 (.45)	.29 (.46)
Sulks	.36 _a (.48)	.46 _b (.50)	.49 _a (.50)	.36 _b (.48)	.42 _{ab} (.49)
Teases	.55 _a (.50)	.39 _b (.49)	.57 _a (.50)	.40 _b (.49)	.49 _{ab} (.50)
Temper tantrums	.58 _a (.49)	.42 _b (.49)	.14 (.35)	.47 (.50)	.51 (.50)
Threatens	.12 (.32)	.08 (.28)	.33 (.47)	.08 (.28)	.10 (.30)
Loud	.41 (.49)	.41 (.49)	.30 _a (.46)	.45 _b (.05)	.41 _{ab} (.49)
Attention Problems					
Acts too young	.21 (.41)	.25 (.43)	.29 (.46)	.21 (.41)	.20 (.40)
Can't concentrate	.34 (.47)	.32 (.47)	.43 _a (.50)	.31 _b (.46)	.27 _b (.45)
Can't sit still	.60 (.49)	.60 (.49)	.67 (.47)	.57 (.50)	.60 (.49)
Confused	.35 (.48)	.28 (.45)	.45 _a (.50)	.27 _b (.45)	.27 _a (.45)
Daydreams	.02 _a (.15)	.31 _b (.46)	.18 (.38)	.16 (.36)	.11 (.32)
Impulsive	.45 (.50)	.41 (.49)	.48 (.50)	.44 (.50)	.35 (.48)
Poor schoolwork	.32 (.47)	.25 (.43)	.37 _a (.48)	.25 _b (.44)	.27 _{ab} (.45)
Stares blankly	.20 (.40)	.17 (.38)	.23 _a (.42)	.19 _{ab} (.39)	.13 _b (.34)

Note. Means for items that have differing subscripts differ significantly at $p < .01$.

Table 2

Unstandardized Item Loadings in Initial Primary Factor Model by Gender and Religioethnicity.

Items	Gender		Religioethnicity		
	Males (n = 623)	Females (n = 523)	Creoles (n = 286)	Hindus (n = 605)	Muslims (n = 255)
	Anxious/Depressed			Anxious/Depressed	
Cries	.62	.95	.77	.79	.74
Fears school	.71	.62	.53	.69	.93
Fears doing bad	.22	.34	.49	.30	.13
Must be perfect	-.01	.02	.04	-.03	-.06
Feels unloved	.81	.85	.94	.80	.52
Feels worthless	.83	.88	.90	.99	.59
Nervous	1.00=	1.00=	1.00=	1.00=	1.00=
Anxious	.77	.88	.86	.90	.62
Feels guilty	-.04	.07	-.01	.04	-.03
Self-conscious	.66	.72	.69	.66	.75
Talks about suicide	.80	.88	.84	.76	1.10
Worries	.71	.95	.77	.94	.70
	Withdrawn/Depressed			Withdrawn/Depressed	
Rather be alone	.43	.69	.32	.52	.70
Won't talk	.30	.75	.69	.42	.27
Secretive	.79	.81	.71	.70	.74
Shy	.23	.52	.40	.38	.16
Underactive	.51	.67	.56	.61	.61
Unhappy	1.00=	1.00=	1.00=	1.00=	1.00=
Withdrawn	.64	1.00	.67	.75	1.02
	Somatic Complaints			Somatic Complaints	
Nightmares	.71	.63	.61	.65	.75
Dizzy	1.00=	1.00=	1.00=	1.00=	1.00=
Overtired	.77	.67	.71	.64	.71
Aches	.68	.71	.62	.68	.73
Headaches	.68	.70	.65	.73	.64
Nausea	.73	.54	.33	.90	.40
Stomachaches	.25	.35	.20	.42	.15
Vomits	.63	.44	.51	.51	.59
	Rule-Breaking Behavior			Aggressive/Rule-Breaking Behavior	
Bad friends	.45	.76	.90	.65	.62
Lies/cheats	1.00=	1.00=	1.00=	1.00=	1.00=
Runs away	.81	.93	1.00	1.00	.73
Steals	.69	.30	.44	.42	.72
Swears	.76	.97	.68	1.08	.81
	Aggressive Behavior				
Argues	.66	.68	.99	.85	.65

Items	Gender		Religioethnicity		
	Males (<i>n</i> = 623)	Females (<i>n</i> = 523)	Creoles (<i>n</i> = 286)	Hindus (<i>n</i> = 605)	Muslims (<i>n</i> = 255)
Cruel to others	.89	.88	1.30	1.15	1.07
Demands attention	.53	.44	.84	.62	.48
Destroys own things	.77	.80	1.22	.87	.97
Destroys others' things	.86	.94	1.42	.84	1.23
Disobedient at home	.92	.97	1.22	1.20	1.16
Disobedient at school	.52	.66	.41	.84	1.21
Fights	.62	.80	.99	1.01	.76
Hits people	.75	.63	.96	.85	1.00
Screams	.83	.73	1.07	.97	1.07
Stubborn	1.00=	1.00=	1.34	1.25	1.32
Moody	.68	.86	1.16	.98	.92
Sulks	.67	.77	1.10	.82	.85
Teases	.79	.74	.99	1.00	.88
Temper tantrums	.69	.81	1.00	1.00	.97
Threatens	.79	.69	1.16	.88	.04
Loud	-.09	.09	.15	.08	-.12
	Attention Problems			Attention Problems	
Acts too young	.43	.18	.34	.39	-.04
Can't concentrate	.55	.53	.49	.59	.44
Can't sit still	.40	.56	.56	.48	.38
Confused	1.00=	1.00=	1.00=	1.00=	1.00=
Daydreams	.58	.77	.74	.57	.64
Impulsive	.66	.46	.64	.61	.55
Poor schoolwork	.63	.55	.48	.67	.55
Stares blankly	.82	.96	.95	1.00	.77

Note. Reference indicators fixed at 1.0 are indicated by "1.00 =" in cell.

Table 3

Item Loadings, Factor Means, and Factor Reliabilities of the Final Primary Factor Models.

	Gender			Religioethnicity			
	Configural		Metric	Configural			Metric
	Boys (<i>n</i> = 623)	Girls (<i>n</i> = 523)	Boys/girls (<i>n</i> = 1,146)	Creole (<i>n</i> = 286)	Hindu (<i>n</i> = 605)	Muslim (<i>n</i> = 255)	Creole/Hindu/Muslim (<i>n</i> = 1,146)
Anxious/Depressed							
Fears school	.48	.42	.45	.37	.40	.67	.45
Fears doing bad	.18	.27	.22	.33	.25	.13	.24
Feels unloved	.59	.58	.58	.59	.54	.35	.53
Feels worthless	.63	.62	.62	.58	.72	.42	.62
Nervous	.71	.66	.68	.65	.62	.78	.67
Anxious	.57	.58	.58	.51	.61	.48	.56
Self-conscious	.49	.48	.48	.42	.46	.59	.47
Latent factor mean	0	0	0/.05	0	0	0	0 _a /.67 _b /.76 _b
Coefficient <i>H</i>	.77	.75	.76	.72	.76	.78	.74
Withdrawn/Depressed							
Rather be alone	.50	.64	.59	.36	.59	.63	.56
Won't talk	.39	.62	.53	.72	.41	.31	.50
Underactive	.51	.52	.51	.56	.50	.48	.50
Withdrawn	.78	.99	.92	.71	.93	.96	.90
Latent factor mean	0	0	0/.09	0	0	0	0 _a /.38 _b /.35 _b
Coefficient <i>H</i>	.70	.98	.87	.73	.88	.93	.84
Somatic Complaints							
Nightmares	.56	.56	.56	.52	.53	.64	.55
Dizzy	.79	.85	.82	.86	.81	.79	.82
Overtired	.57	.55	.56	.58	.48	.54	.53
Aches	.52	.63	.58	.51	.57	.56	.56
Headaches	.51	.61	.56	.57	.56	.48	.55
Stomachaches	.18	.30	.25	.20	.25	.11	.21
Vomits	.46	.41	.44	.46	.36	.55	.43
Latent factor mean	0	0	0 _a /.21 _b	0	0	0	0 _a /.65 _b /.87 _b
Coefficient <i>H</i>	.77	.83	.81	.83	.79	.80	.80
Aggressive Behavior							
Destroys own things	.60	.67	.63	.72	.51	.63	.62
Destroys others' things	.71	.81	.75	.86	.52	.80	.75
Disobedient at home	.68	.73	.71	.62	.70	.72	.69
Hits people	.53	.50	.52	.53	.42	.66	.53
Screams	.63	.52	.58	.54	.56	.63	.57
Stubborn	.78	.82	.80	.69	.80	.88	.79
Teases	.62	.58	.60	.57	.59	.55	.58
Threatens	.59	.58	.59	.64	.52	.69	.60

	Gender			Religioethnicity			
	Configural		Metric	Configural			Metric
	Boys (<i>n</i> = 623)	Girls (<i>n</i> = 523)	Boys/girls (<i>n</i> = 1,146)	Creole (<i>n</i> = 286)	Hindu (<i>n</i> = 605)	Muslim (<i>n</i> = 255)	Creole/Hindu/Muslim (<i>n</i> = 1,146)
Latent factor mean	0	0	0 _a /-.32 _b	0	0	0	0 _a /-.76 _b /-.53 _b
Coefficient <i>H</i>	.86	.88	.87	.88	.83	.91	.87
Attention Problems							
Acts too young	.35	.16	.26	.29	.27	.05	.23
Can't concentrate	.42	.46	.44	.41	.41	.39	.41
Confused	.78	.79	.79	.76	.69	.88	.77
Impulsive	.49	.38	.44	.48	.43	.48	.45
Poor schoolwork	.48	.46	.47	.37	.50	.47	.46
Stares blankly	.62	.69	.65	.66	.69	.62	.66
Latent Factor Mean	0	0	0 _a /-.20 _b	0	0	0	0 _a /-.53 _b /-.70 _b
Coefficient <i>H</i>	.76	.77	.76	.74	.73	.83	.75

Note. Loadings reflect UVI identification. Abbreviations for models are as follows: Configural = configural invariant model; Metric = metric invariant model. Subscripts mark latent means that differ significantly among groups. Alpha levels are all $p < .001$, except for the lower mean for Muslim as compared with Creole on the Withdrawn/Depressed factor, $p < .01$. Mean comparisons between Hindu and Muslim, ascertained through the use of an alternative reference group, are presented in text.

Table 4

Item Loadings, Factor Means, and Factor Reliabilities of the Final Secondary Factor Models.

	Gender			Religioethnicity			
	Configural		Metric	Configural			Metric
	Boys (<i>n</i> = 623)	Girls (<i>n</i> = 523)	Boys/ girls (<i>n</i> = 1,146)	Creole (<i>n</i> = 286)	Hindu (<i>n</i> = 605)	Muslim (<i>n</i> = 255)	Creole/Hindu/Muslim (<i>n</i> = 1,146)
Internalizing							
Nervous	.70	.64	.67	.65	.61	.74	.66
Fears school	.47	.42	.45	.37	.40	.64	.45
Fears doing bad	.18	.27	.22	.33	.24	.13	.24
Feels unloved	.58	.56	.58	.59	.54	.33	.53
Feels worthless	.62	.62	.62	.59	.72	.40	.61
Anxious	.58	.58	.58	.52	.62	.47	.56
Self-conscious	.48	.48	.48	.42	.46	.57	.47
Rather be alone	.23	.34	.28	.11	.27	.35	.24
Won't talk	.15	.38	.27	.38	.18	.07	.23
Underactive	.30	.35	.32	.35	.33	.30	.32
Withdrawn	.38	.53	.45	.39	.41	.50	.42
Dizzy	.75	.82	.79	.81	.77	.74	.78
Nightmares	.54	.54	.54	.50	.50	.62	.53
Overtired	.55	.53	.54	.56	.46	.51	.51
Aches	.51	.60	.56	.48	.55	.54	.53
Headaches	.49	.59	.54	.53	.53	.45	.52
Stomachaches	.16	.28	.23	.17	.22	.11	.19
Vomits	.45	.40	.42	.43	.34	.53	.41
Confused	.70	.72	.71	.66	.64	.80	.69
Can't concentrate	.37	.41	.39	.35	.36	.35	.36
Stares blankly	.58	.64	.61	.61	.65	.57	.61
Latent factor mean	.00	.00	.00/.08	.00	.00	.00	0 _a /.66 _b /.81 _b
Coefficient <i>H</i>	.89	.91	.90	.89	.89	.90	.89
Externalizing							
Stubborn	.76	.81	.78	.68	.79	.86	.78
Destroys own things	.58	.66	.62	.71	.50	.61	.60
Destroys others' things	.68	.80	.74	.84	.51	.79	.74
Disobedient at home	.67	.73	.70	.61	.70	.70	.68
Hits people	.52	.49	.51	.51	.42	.64	.51
Screams	.62	.51	.57	.54	.55	.61	.56
Teases	.60	.56	.59	.56	.58	.54	.57
Threatens	.58	.57	.58	.62	.52	.68	.59
Impulsive	.49	.39	.44	.45	.46	.46	.44
Poor schoolwork	.46	.47	.47	.35	.52	.46	.46
Latent factor mean	.00	.00	0 _a /.32 _b	.00	.00	.00	0 _a /.73 _b /.55 _b

	Gender			Religioethnicity			
	Configural		Metric	Configural			Metric
	Boys (<i>n</i> = 623)	Girls (<i>n</i> = 523)	Boys/ girls (<i>n</i> = 1,146)	Creole (<i>n</i> = 286)	Hindu (<i>n</i> = 605)	Muslim (<i>n</i> = 255)	Creole/Hindu/Muslim (<i>n</i> = 1,1460)
Coefficient <i>H</i>	.86	.88	.87	.88	.85	.90	.87

Note. Loadings reflect UVI identification. Abbreviations for models are as follows: Configural = configural invariant model; Metric = metric invariant model. Subscripts mark latent means that differ significantly. Alpha levels are all $p < .001$. Mean comparisons between Hindu and Muslim, ascertained through the use of an alternative reference group, are presented in text.