

# The Application of the Preschool Child Behavior Checklist and the Caregiver–Teacher Report Form to Mainland Chinese Children: Syndrome Structure, Gender Differences, Country Effects, and Inter-Informant Agreement

Jianghong Liu · Halina Cheng · Patrick W. L. Leung

Published online: 7 September 2010  
© Springer Science+Business Media, LLC 2010

**Abstract** Preschool children have long been a neglected population in the study of psychopathology. The Achenbach System of Empirically Based Assessment (ASEBA), which includes the Child Behavior Checklist/1.5-5 (CBCL/1.5-5) and the Caregiver-Teacher Report Form (C-TRF), constitutes the few available measures to assess preschoolers with an empirically derived taxonomy of preschool psychopathology. However, the utility of the measures and their taxonomy of preschool psychopathology to the Chinese is largely unknown and has not been studied. The present study aimed at testing the cross-cultural factorial validity of the CBCL/1.5-5 and C-TRF, as well as the applicability of the taxonomy of preschool psychopathology they embody, to Mainland Chinese preschoolers. Country effects between our Chinese sample and the original U.S. sample, gender differences, and cross-informant agreement between teachers and parents were also to be examined. A Chinese version of the CBCL/1.5-5 and C-TRF was completed by parents and teachers respectively on 876 preschoolers in Mainland China. Confirmatory factor analysis (CFA) confirmed the original, U.S.-derived second order, multi-factor model best fit the Chinese preschool data of the CBCL/1.5-5 and C-TRF. Rates of total behavior problems in Chinese preschoolers were largely similar to those in American preschoolers. Specifically, Chinese preschoolers scored higher on internalizing problems while American

preschoolers scored higher on externalizing problems. Chinese preschool boys had significantly higher rates of externalizing problems than Chinese preschool girls. Cross-informant agreement between Chinese teachers and parents was relatively low compared to agreement in the original U.S. sample. Results support the generalizability of the taxonomic structure of preschool psychopathology derived in the U.S. to the Chinese, as well as the applicability of the Chinese version of the CBCL/1.5-5 and C-TRF.

**Keywords** CBCL · Externalizing · Internalizing · Preschool · Cross-cultural · Chinese children · CFA · Confirmatory factor analysis · Validation

Childhood behavior problems represent an important topic in developmental psychopathology. Internalizing behavior problems in early childhood are risk factors for teenage and adult depression, anxiety, and suicide, while externalizing behavior problems are risk factors for later juvenile delinquency, adult crime, and violence (Farrington 1989; Moffitt 1993; Raine 2002). Thus, identifying early childhood behavior problems is critically important for understanding and preventing the development of problem behaviors later in life (Liu and Wuerker 2005). However, cross-cultural research on internalizing and externalizing behavior in early childhood (e.g., preschool) is lacking. Given the potential importance of cultural and social factors in psychopathology (Weisz et al. 2006), this study seeks to better understand the generalizability and validity of American constructs on early childhood behavior problems in a non-Western culture, namely Mainland Chinese.

Childhood behavior problems can be measured in a number of different ways. The most popular approach has been to use rating scales that are completed by either parents or teachers. The best example of such an approach

---

J. Liu (✉)  
Schools of Nursing and Medicine, University of Pennsylvania,  
418 Curie Blvd., Room 426, Claire M. Fagin Hall,  
Philadelphia, PA 19104-6096, USA  
e-mail: jhliu@nursing.upenn.edu

H. Cheng · P. W. L. Leung (✉)  
Department of Psychology,  
Chinese University of Hong Kong,  
Hong Kong, China  
e-mail: pleung@cuhk.edu.hk

is the Achenbach System of Empirically Based Assessment (ASEBA), which consists of three parallel questionnaires: the Child Behavior Checklist (CBCL) using parents as informants, the Teacher Report Form (TRF), and the Youth Self-Report (YSR). The ASEBA is based on carefully conducted empirical studies and is designed to assess, in a standardized format, behavioral problems and social competencies (Achenbach and Rescorla 2001). These three rating scales have been translated into 80 languages, with thousands of published empirical studies in over 60 societies supporting their psychometric properties and usefulness in the research of childhood psychopathology (ASEBA 2009). For school-age children and adolescents, evidence of the taxonomic construct validity of ASEBA syndromes has been provided by confirmatory factor analysis (CFA) of the CBCL from 30 countries ( $N=58,043$ ), of the TRF from 20 countries ( $N=30,030$ ), and of the YSR from 23 countries ( $N=30,243$ ) (Ivanova et al. 2007a, b, c), including Chinese communities in Mainland China, Taiwan, and Hong Kong. Particularly, the test-retest reliability and criterion validity of the Chinese versions of these three rating scales have been established (Leung et al. 2006). For preschool-age children, a similar strategy was applied to systematically derive an empirically based taxonomy of preschool problem behaviors, resulting in development of preschool versions of the CBCL and the Caregiver-Teacher Report Form (C-TRF). The current versions of the CBCL/1.5-5 and C-TRF were published in 2000 in English (Achenbach and Rescorla 2000).

The development of the preschool versions of the CBCL and C-TRF offers several notable merits. First, it fills in the gaps in our knowledge on preschool psychopathology and constructs a taxonomy of preschool behavior problems by identifying syndromes and higher-order broad-band problems. Second, the preschool versions explicitly address the developmental concerns of children during this age period. While it shares some items with the school-age version, there are specific items written for the preschool period, which consequently form different syndromes, including aggression, defiance, hyperactivity, fears, and social anxiety. These emotions and behaviors are all common in preschoolers (e.g., Tremblay 2004).

Factor analysis performed on the CBCL/1.5-5 and C-TRF has separately produced six syndromes for each measure: Emotionally Reactive, Anxious/Depressed, Aggressive Behavior, Attention Problems, Somatic Complaints, and Withdrawn. The CBCL/1.5-5 has a seventh syndrome, Sleep Problems, which consists of sleep-related items that are not assessed in the C-TRF. When these preschool syndromes are submitted to a second-order factor analysis, two broad-band problems, labeled “Internalizing” and “Externalizing,” emerge. The former includes the syndromes of Emotionally Reactive, Anxious/Depressed, Somatic Complaints, and

Withdrawn, while the latter includes syndromes of Aggressive Behavior and Attention Problems. The syndrome of Sleep Problems in CBCL/1.5-5 stands alone. In addition to those empirically derived syndrome scales, the CBCL/1.5-5 and C-TRF also provide five DSM-Oriented Scales (i.e., Affective, Anxiety, Somatic, Attention-Deficit/Hyperactivity, Oppositional Defiant, and Pervasive Developmental Scales), which are not empirically-derived by factor analysis but are constructed via expert clinical consensus based on the criteria of the DSM-IV (APA 1994).

There are several gaps in the literature in regards to preschool psychopathology. First, few instruments are available to study behavior problems in preschool children. Research in this population has been lagging behind research conducted on older children and adolescents for at least 30 years (Egger and Angold 2006). Second, as noted above, cross-cultural research on preschool child behavior is rarely conducted. Third, empirical evidence regarding the cross-cultural validity of childhood diagnostic criteria and classification is still largely lacking (Canion and Alegria 2008). To date, the taxonomy of preschool psychopathology, exemplified in the CBCL/1.5-5 and C-TRF, has rarely been evaluated in a non-Western culture, such as a Chinese culture. Only one prior study examined the taxonomy of preschool psychopathology in the CBCL/1.5-5 in a sample of Chinese girls adopted from Mainland China to the United States (U.S.) (Tan et al. 2007). However, despite the fact that Chinese girls are involved in this study, the rating of the CBCL/1.5-5 is in fact done by American parents on Chinese girls living in the U.S. This study cannot be strictly considered as cross-cultural; a real test of the cross-cultural validity of CBCL/1.5-5 and C-TRF in a Chinese sample should be conducted with Chinese parents/teachers and preschoolers living on Chinese soil. To the best knowledge of the authors, this is the first study of this kind.

Empirical evidence concerning the cross-cultural validity of the taxonomy of problem behaviors in CBCL and its parallels can only be found from research conducted with school-aged children. In a major review of Western measures and East Asian populations, Leung and Wong (2003) concluded that there was no empirical support for the need of culture-specific diagnostic constructs for Asian children/adolescents. Two studies, one with Taiwanese participants and another with Thai participants, found that when culture-specific problem items were added to the CBCL and TRF, these items did not produce a new factor or diagnostic construct in factor analyses. Instead, they loaded onto different existing factors/constructs, alongside original items belonging to the CBCL or TRF (Cederblad et al. 2001; Yang et al. 2000). However, Weisz and his colleagues in Thailand (2006) found that many of the CBCL syndromes did not demonstrate cross-cultural

factorial validity in Thailand. Instead, some new diagnostic constructs emerged, such as delayed maturation, indirect aggression, and/or delinquency. Yet, as mentioned above, the factor structure (i.e., taxonomy of problem behaviors) of the CBCL and its two parallels, the TRF and YSR, had recently been confirmed in more than 100,000 children/adolescents from 20 to 30 diverse societies (Ivanova et al. 2007a, b, d). However, it must be noted that in this latest round of world-wide analyses of the factor structure of CBCL and its parallels, they used advanced statistical techniques—Weighted Least Squares (WLS) and robust standard error with mean- and variance-adjusted fit statistics (WLSMV), an asymptotically distribution-free (ADF) estimator—to handle the ordinal response scale of the CBCL/TRF/YSR, and their non-normal distributions. We would also adopt this advanced statistical approach in this study in order to address the ordinal scale and non-normal distributed data of CBCL/1.5-5 and C-TRF.

Given the discrepant findings and particularly, the lack of cross-cultural data in preschoolers, we first propose to independently assess Achenbach's and Rescorla's (2000) U.S.-derived taxonomy of preschool psychopathology in a different culture/country outside the U.S., in this case, Chinese culture/Mainland China, using both the CBCL/1.5-5 and the C-TRF in boys and girls. In view of the success of the school-age CBCL/TRF/YSR in confirming factorial validity in a huge sample across numerous societies, we hypothesize in Chinese preschoolers that the original multi-factor model will be superior to a one-factor model and that the two higher-order broad-band and six associated syndromes model of preschool psychopathology identified by Achenbach and Rescorla (2000) will adequately represent the factorial structure of the CBCL/1.5-5 and the C-TRF.

Second, most previous studies of both school-age and preschool populations (Ivanova et al. 2007a, b, c; Tan et al. 2007) have conducted factor analysis for the individual syndromes and for the broad-band problems separately, or for the former only. This study will attempt to test both syndromes and broad-band problems together in one CFA. This will represent a more complete test of the hierarchical arrangement of the syndromes and broad-band problems that constitute a taxonomy of preschool psychopathology.

Third, if the factorial validity of the two preschool measures is confirmed, scores from our Mainland Chinese sample can be compared to those of the original U.S. sample by examining the rates of behavioral and emotional problems among Chinese preschoolers and comparing them to those of the original U.S. sample. Besides country effects, we can also examine gender differences and inter-informant agreement for comparison to findings from the original U.S. sample. Recent studies of the school-age CBCL, TRF, and YSR have reported a fair degree of consistency between total problem scores as well as gender differences across a large number (21 to 31) of very diverse

cultures/societies (Rescorla 2007; Rescorla et al. 2007a, b). In view of the above findings, we hypothesize that in Chinese preschoolers, modest cross-informant agreement between teachers and parents will emerge. We further predict that significant gender differences will be found in both the CBCL and C-TRF, with boys scoring higher on externalizing behavior and girls scoring higher on internalizing behavior.

## Methods

### Participants and Procedures

The current study was part of a larger population-based community cohort study of 1,656 Chinese children (55.5% boys, 44.5% girls) initially recruited between the fall of 2004 and the spring of 2005 from four preschools in the town of Jintan, located in the southeastern coastal region of Mainland China. In China, preschools are called kindergartens and enroll children from ages 3–7, after which children enter the elementary school system. Preschools are divided into junior (3–4 years old), middle (4–5 years old), and senior levels (5+ years old). The four we selected were representative of the geographic, social, and economic profile of all preschools in Jintan. Detailed sampling and research procedures of this larger cohort study have been described elsewhere (Liu et al. 2010). Briefly, all children and parents taking part in the original cohort study were invited to participate for assessment of children's behaviors while the children were in the final few months of their senior year in preschool (spring 2005 to spring 2007). The response participation rate was 97%.

Parents and teachers were asked to assess the children with the Chinese version of the CBCL/1.5-5 and C-TRF. The Chinese version had been translated and back-translated and subsequently finalized by consensus amongst a team of experienced child clinical psychologists in order to ensure adequacy of the translation. Following the standard procedure of discarding questionnaires with missing data for more than eight items, the final data set was comprised of 1,209 CBCL/1.5-5 and C-TRF (54.7% boys, 45.2% girls) responses. Each teacher completed C-TRF ratings on approximately 40 children. In order to avoid overburdening the teachers in a short period of time, they were given the forms during the middle of the semester and were collected during the summer school break. They were instructed to spread out their rating of students throughout the semester and the summer holiday so that there was sufficient time to rate each of the 40 students.

Since some of the children were beyond the age limit of the CBCL/1.5-5 and C-TRF (i.e., over 71 months), the current analysis only addressed the subset of the original

sample that was under age 6 to adhere to the age requirement of the measures. Our final data set for analysis was thus comprised of 876 preschoolers (52.7% boys,  $N=462$ , 47.3% girls,  $N=414$ ). The mean age of this subset was 66.6 months ( $SD=5$ , range=50–71). We are aware that this final sample represents the upper age range of the instruments, which were designed for ages 18 to 71 months.

### Data Analysis

We performed CFA on the 67 CBCL/1.5-5 items that loaded significantly on the seven-syndrome and two-broad-band-problem factor structure derived from Achenbach and Rescorla's (2000) original U.S. sample. The same procedure was performed on the 66 C-TRF items that loaded significantly on the six-syndrome and two-broad-band-problem factor structure derived from the same U.S. sample. Following Achenbach and Rescorla's (2000) procedures, we dichotomized the data of both the CBCL/1.5-5 and C-TRF by converting item scores to 0 versus 1 or 2 to "avoid statistical risk associated with low frequency cells." The tetrachoric correlation matrices were analyzed with the WLSMV estimation method in Mplus 5.0 (Muthen and Muthen 1998–2007). Each item was specified to only load on the one syndrome factor that it was supposed to measure; that is, no cross-loadings were allowed and error-covariances were fixed to zero. For the one-factor models, all 67/66 items of the CBCL/C-TRF were used to measure a single underlying variable, with uncorrelated item residuals. CFA was not performed on the five DSM-Oriented Scales because they were not empirically derived by factor analysis. However, we did examine cross-cultural and gender differences, as well as cross-informant agreement, on these scales.

We then ran the one-level one-factor models for the CBCL and C-TRF. We compared the model fit with the above CFA that tested the factor structures derived from the original U.S. sample. Although we report most conventional fit indices, chi-square, and degrees of freedom, RMSEA was chosen to be the major indicator for model fit because we employed categorical indicators in the model and used WLSMV estimator methods in fitting the CFA. The RMSEA had been identified as the best performing index for WLSMV, with values  $\leq 0.06$  indicating good fit (Yu and Muthen 2002) and  $\leq 0.08$  indicating acceptable fit (Browne and Cudeck 1993). 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 and Lewis 1973). These statistics, however, are considered to be secondary to those of the RMSEA because it is still unknown whether the CFI and TLI are appropriate for use with categorical data. Hu and Bentler (1999) proposed that CFI and TLI values  $>0.95$  be required for good model fit. However, this criterion has been criticized for being too

stringent and for often incorrectly rejecting properly defined complex models (Marsh et al. 2004). Given the complexity of our model, we used Browne and Cudeck's (1993) less stringent criteria of  $>0.90$  for good fit and 0.80 to 0.90 for acceptable fit. Apart from statistical criteria, other substantive concerns, such as the interpretability of the latent factors, were also considered when selecting models. T-tests were carried out to evaluate the effects of gender on the derived syndromes. Given the multiple tests (15) performed, the Bonferroni-corrected p value was set at  $\leq 0.003$ . We used independent sample t-tests (Bonferroni-corrected) to compare the rates of problems reported on the syndromes from the Chinese sample with those from the U.S. sample. The degree of agreement between parents and teachers on the rated problems was assessed by Pearson correlations and paired t-tests (Bonferroni-corrected) on the syndrome scores of the CBCL/1.5-5 and C-TRF. Lastly, we compared the degree of cross-informant agreement between the Chinese and U.S. samples with Fisher's z test.

### Results

Missing data considerations were made for the CFA models. There were few cases missing on the checklist, and the missing cases varied from 0 to 33 for the parent sample and 0 to 4 for the teacher evaluations. All items were equally likely to be skipped. The missing data imputation function in the M-Plus software was used to retain all cases (Listwise deletion would result in a 23% and 4% of data loss for the CBCL and C-TRF model respectively, and the results are very similar to the ones with imputation, reported below).

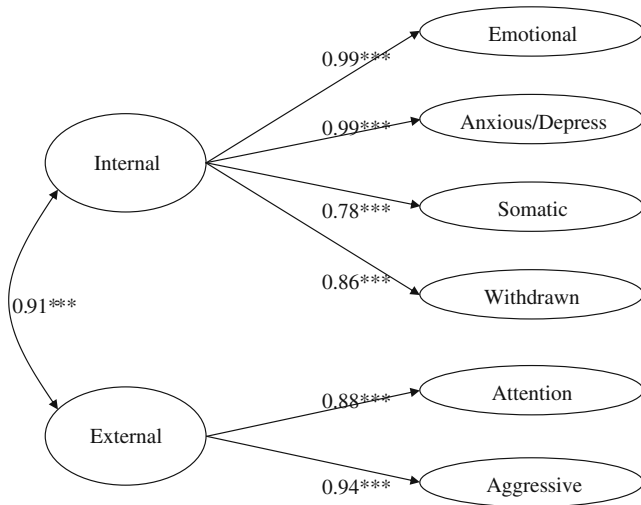
*CFA Models' Fit* The one-factor one-level CBCL model reported a chi-square of 1186.7 ( $\chi^2$ ), with 415 (d.f.)° of freedom, CFI=0.887 (CFI), TLI=0.948 (TLI), and RMSEA=0.046 (RMSEA). The second order multifactor CBCL model reported a chi-square of 964.6, with 416° of freedom, CFI=0.920, TLI=0.963, and RMSEA=0.039. The one-factor one-level C-TRF model reported a chi-square of 1861.2 ( $\chi^2$ ), with 213° of freedom, CFI=0.688, TLI=0.798, and RMSEA=0.094. The second order multifactor C-TRF model reported a chi-square of 1342.2, with 216° of freedom, CFI=0.787, TLI=0.864, and RMSEA=0.077.

Based on the criteria of RMSEA  $\leq 0.06$  indicating a good fit (Yu and Muthen 2002) and  $\leq 0.08$  an acceptable fit (Browne and Cudeck 1993), both CBCL models attained a good fit, and the second order C-TRF model had an acceptable fit. The one-level C-TRF model did not reach an acceptable level of model fit. Thus, for the C-TRF, the second order six-syndrome model was clearly superior to the one-level one-factor model. We compared the change in

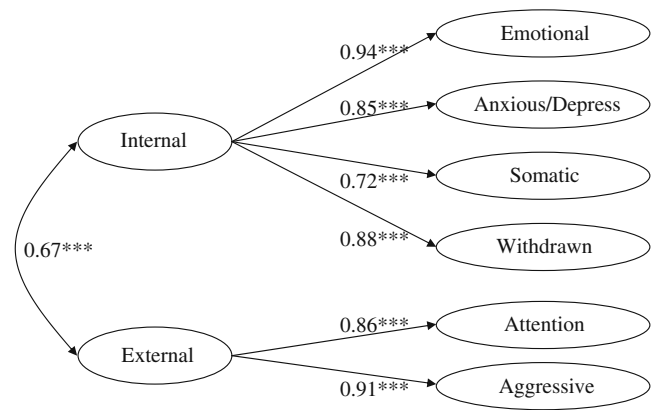
RMSEA from a one-factor model to a two-level model. The difference for the CBCL models was 0.007, while it was 0.017 for the C-TRF model. Both times, the second order multi-factor models had a smaller RMSEA, indicating a better fit than the respective one-level one-factor models. We considered the original multi-factor models from Achenbach and Rescorla (2000) preferable to the one-factor models. The case for the original multi-factor model of the CBCL/1.5-5 was less obvious, given that the RMSEA for the one-factor model was only slightly worse ( $\Delta$ RMSEA=0.007).

The above analyses were performed with the full sample of boys and girls. We repeated the analyses by gender, testing our preferred second order, multi-factor models of the CBCL/1.5-5 and C-TRF. The results were largely similar to those with the full sample (CBCL RMSEA= 0.043 & 0.044 for boys & girls respectively; C-TRF RMSEA=0.080 & 0.063 for boys and girls respectively). This supported our preference for the second order, multi-factor models, which were applicable across the two genders and questionnaires.

Figures 1 and 2 show the factor loadings of the two-level CBCL and C-TRF models. Correlations between the two broadband Internalizing and Externalizing Problems on the CBCL/1.5-5 and C-TRF were 0.91 and 0.67 respectively. For both instruments, all relevant derived syndromes had significant ( $p \leq 0.01$ ), positive, and substantial loadings on their respective higher-order factors, Internalizing and Externalizing Problems. The loadings on the Internalizing factor for the CBCL/1.5-5 and C-TRF were respectively: Emotionally Reactive 0.99/0.94; Anxious/Depressed 0.99/0.85; Somatic Complaints 0.78/0.72; and Withdrawn 0.86/0.88. For the Externalizing factor, loadings on the CBCL/1.5-5



**Fig. 1** Second order CFA for CBCL/1.5-5 in a Chinese Sample (The Sleep Problems Syndrome factor is only present in CBCL and not in C-TRF. Therefore, although this factor was included in the analysis, it is not shown in the figure above)



**Fig. 2** Second order CFA for C-TRF in a Chinese sample

and C-TRF were respectively: Attention Problems 0.88/0.86 and Aggressive Behavior 0.94/0.91. The second-order factor loadings found in the Chinese sample were relatively higher than those reported in the U.S. sample (Achenbach and Rescorla 2000). The latter were mean loadings, obtained by averaging the loadings across four second-order analyses (by gender and by questionnaire). For Internalizing factor, they were: Emotionally Reactive 0.81; Anxious/Depressed 0.61; Somatic Complaints 0.60; and Withdrawn 0.54, while for Externalizing factor, they were: Attention Problems 0.67 and Aggressive Behavior 0.75 (Achenbach and Rescorla 2000).

**Latent Factor Correlations** All latent factor correlations among the syndromes on both the CBCL/1.5-5 and C-TRF were positive and statistically significant ( $p \leq 0.01$ ). These were disattenuated correlations, which have measurement errors controlled. The latent factor correlation coefficients among the seven syndromes of the CBCL/1.5-5 had a median of 0.78 (range=0.64–0.99), while those among the six syndromes of the C-TRF had a median of 0.60 (range=0.36–0.98). Details of the latent factor correlations among the CBCL and C-TRF syndromes are presented in Table 1. Since all correlations on both questionnaires were below unity and the majority of them, with the exception of Anxious/Depressed with Emotionally Reactive in CBCL/1.5-5 and C-TRF, had even their upper limits (using a stringent 99.7% confidence interval) below unity, such results were supportive of a correlated multi-factor model for both questionnaires, in addition to a smaller RMSEA, mentioned above.

**Item Factor Loadings** CFA results indicated that for both questionnaires, all relevant items had significant ( $p \leq 0.01$ ), positive, and substantial loadings on the syndromes they were supposed to belong to and measure. For the CBCL/1.5-5, the median and mean item loadings were 0.64 and 0.63 respectively (range=0.33–0.84), which were largely comparable to the 0.54 and 0.55 (range=0.16–0.96) reported in the

**Table 1** Factor correlations (disattenuated) among the CBCL syndromes (shown above diagonal) and C-TRF syndromes (shown below diagonal)

	Anxious/ depressed	Emotionally reactive	Somatic complaints	Withdrawn	Sleep problems	Attention problems	Aggressive behavior
Anxious/Depressed		0.99	0.75	0.82	0.79	0.78	0.89
Emotionally Reactive	0.98		0.81	0.89	0.80	0.78	0.82
Somatic Complaints	0.70	0.68		0.66	0.74	0.65	0.64
Withdrawn	0.65	0.80	0.51		0.74	0.69	0.73
Sleep Problems	NA	NA	NA	NA		0.71	0.72
Attention Problems	0.48	0.36	0.36	0.60	NA		0.83
Aggressive Behavior	0.61	0.42	0.48	0.54	NA	0.78	

original U.S. sample (Achenbach and Rescorla 2000). For the C-TRF, the median and mean loadings were 0.70 and 0.69 respectively (range 0.28–0.88), comparable to 0.71 and 0.66 (range 0.31–0.94) reported in the U.S. sample. (A table detailing the first-level item factor loadings of our Chinese sample, as compared to those of the original US sample, can be obtained from the corresponding author.)

*Correlations Among Syndromes and Broadband Problems* In order to compare the magnitudes of correlations among syndromes and broadband problems reported for our Chinese sample and the original U.S. sample, Pearson correlations were computed. We used listwise deletion to deal with missing cases so that only those who had all items reported would be included in the analysis. All correlations were positive and statistically significant ( $p \leq 0.01$ ). The correlations among the seven syndromes of the CBCL/1.5-5 had a median and a mean of 0.51 and 0.53 respectively (range 0.38–0.71), which were largely comparable to those obtained in the original U.S. sample (the median and the mean being identical at 0.39, range 0.17–0.67) (Achenbach and Rescorla 2000). The correlations among the six syndromes of the C-TRF in our Chinese sample had a median and a mean of 0.44 and 0.42 respectively (range 0.22–0.67), values which were close to the 0.41 and 0.40 (range 0.09–0.68) reported in the U.S. sample. Correlations between the two broadband Internalizing and Externalizing Problems in our CBCL/1.5-5 and C-TRF were 0.75 and 0.49 respectively, compared to 0.59 and 0.62 respectively reported in the U.S. sample. According to the Fisher's  $z$  test, in our Chinese sample, Internalizing and Externalizing Problems were significantly more highly correlated in the CBCL/1.5-5 than in the C-TRF (0.75 vs. 0.49,  $z=8.03$ ,  $p < 0.01$ ). However, this was not the case for the U.S. sample (0.59 vs. 0.62,  $z=-0.99$ ,  $p > 0.05$ ). Across samples, in the CBCL/1.5-5, our Chinese sample reported a higher correlation between Internalizing and Externalizing Problems than the U.S. sample (0.75 vs. 0.59,  $z=5.07$ ,  $p < 0.01$ ), while the reverse was true for the C-TRF (0.49 vs. 0.62,  $z=-4.19$ ,  $p < 0.01$ ). The significantly higher correlation coefficient

between Internalizing and Externalizing Problems reported by parents in the Chinese sample had already been reflected by the CFA results in which, compared to the C-TRF, the multi-factor model in the CBCL/1.5-5 was less differentiated from the one-factor model.

*Effects of Gender on the CBCL and C-TRF* While American boys scored significantly higher than American girls only on the CBCL/1.5-5 DSM-Oriented Scale of Attention Deficit/Hyperactivity Problems (Achenbach and Rescorla 2000), Chinese boys scored significantly ( $p \leq 0.003$ ) higher than Chinese girls on scales of Attention Problems, Aggressive Behavior, Externalizing, and Oppositional Defiant Problems, with effect sizes (Cohen's  $d$ ) ranging from 0.20 to 0.36 (mean=0.29) (see Table 2). Despite these effects being considered small (Cohen 1988), these gender differences were statistically significant.

With respect to the C-TRF, the Chinese boys also demonstrated significantly higher scores than girls on scales related to Total Problems, Externalizing Problems, Attention Problems, Aggressive Behavior, Attention Deficit/Hyperactivity Problems, and Oppositional Defiant Problems with small-to-medium effect sizes (mean=0.46, range=0.31 to 0.54), similar to findings in the original U.S. sample (Achenbach and Rescorla 2000) (see Table 3). In contrast, Chinese girls yielded significantly higher scores on Anxious/Depressed, a finding not reported in the U.S. sample.

*Comparison of Scores Between Chinese and U.S. Samples* Independent sample  $t$ -tests were carried out to assess whether Chinese and U.S. samples differed significantly on their scores on the CBCL/1.5-5 (Table 4). With a few exceptions (e.g., Emotional Reactive, Total Problems, Sleep Problems, and Attention Deficit/Hyperactivity Problems), significant differences were found on most scales, with small effect sizes (mean=0.33, range=0.15 to 0.46). The Chinese sample demonstrated significantly higher scores on Internalizing Problems, Anxious/Depressed, Somatic Complaints, Withdrawn, Attention Problems, Affective Problems, Anxiety Problems, and Pervasive Developmental Problems, while the

**Table 2** Mean score comparisons and effect sizes between Chinese boys and girls on the CBCL/1.5-5

	Boys		Girls		t-value	d'
	Mean	SD	Mean	SD		
<b>Syndromes</b>						
Emotionally Reactive	2.76	2.48	2.59	2.37	1.04	0.07
Anxious/Depressed	3.33	2.20	3.24	2.33	0.61	0.04
Somatic Complaints	2.67	2.28	2.66	2.44	0.11	0.00
Withdrawn	2.31	2.30	2.15	2.46	1.00	0.07
Sleep Problems	2.87	2.06	2.73	2.18	0.97	0.07
Attention Problems	3.01	1.91	2.54	1.76	3.70*	0.26
Aggressive Behavior	8.56	5.85	6.63	5.30	4.92*	0.35
Internalizing	10.74	7.34	10.18	7.87	0.99	0.07
Externalizing	11.58	7.24	9.08	6.51	5.14*	0.36
Total Problems	35.75	21.02	31.84	20.69	2.77	0.19
<b>DSM-oriented scales</b>						
Affective Problems	3.33	2.57	3.03	2.75	1.64	0.11
Anxiety Problems	4.41	2.67	4.25	2.72	0.85	0.06
Pervasive Developmental Problems	3.70	3.25	3.54	3.30	0.71	0.05
Attention Deficit/Hyperactivity Problems	5.10	3.33	4.59	2.67	2.50	0.17
Oppositional Defiant Problems	2.92	2.40	2.45	2.35	2.95*	0.20

\* $p \leq 0.003$

U.S. sample had higher scores on Externalizing Problems, Aggressive Behavior, and Oppositional Defiant Problems.

Since significant gender differences were found in the C-TRF in the original U.S. sample, separate norms were

reported for boys and girls. Thus, independent sample t-tests to compare scores of the Chinese sample and the U.S. sample were conducted separately by gender (Table 5). Overall, whenever there was a significant difference on scores between

**Table 3** Mean score comparisons and effect sizes between Chinese boys and girls on the C-TRF

	Boys		Girls		t	d'
	Mean	SD	Mean	SD		
<b>Syndromes</b>						
Emotionally Reactive	1.13	1.52	1.13	1.55	-0.25	0.00
Anxious/Depressed	1.70	1.86	2.12	2.11	-3.13*	0.21
Somatic Complaints	0.55	1.03	0.57	1.02	-0.29	0.02
Withdrawn	2.38	2.71	2.15	2.64	1.29	0.09
Attention Problems	4.01	3.18	2.71	2.85	6.37*	0.43
Aggressive Behavior	6.40	7.11	3.22	4.51	7.94*	0.53
Internalizing	5.73	5.57	5.94	5.91	-0.53	0.04
Externalizing	10.39	9.52	5.93	6.73	8.02*	0.54
Total Problems	22.31	17.91	17.18	15.18	4.58*	0.31
<b>DSM-oriented scales</b>						
Affective Problems	1.18	1.66	1.32	1.83	-1.15	0.08
Anxiety Problems	1.24	1.80	1.41	1.69	-1.45	0.10
Pervasive Developmental Problems	3.02	3.09	2.79	2.91	1.10	0.08
Attention Deficit/Hyperactivity Problems	5.53	4.39	3.79	3.78	6.30*	0.42
Oppositional Defiant Problems	1.30	1.90	0.78	1.43	4.62*	0.31

\* $p \leq 0.003$

**Table 4** Mean score comparisons and effect sizes for Chinese and U.S. children on the CBCL/1.5-5

	Chinese		U.S.		t	d'
	Mean	SD	Mean	SD		
<b>Syndromes</b>						
Emotionally Reactive	2.68	2.43	2.40	2.20	2.34	0.12
Anxious/Depressed	3.29	2.26	2.90	2.30	3.34*	0.17
Somatic Complaints	2.67	2.36	1.80	1.90	7.81*	0.41
Withdrawn	2.24	2.38	1.50	1.70	6.91*	0.36
Sleep Problems	2.80	2.12	2.80	2.40	0.00	0.00
Attention Problems	2.79	1.86	2.50	1.90	3.04*	0.15
Aggressive Behavior	7.64	5.67	10.40	6.40	-8.88*	0.46
Internalizing	10.47	7.60	8.60	6.20	5.09*	0.27
Externalizing	10.39	7.01	12.90	7.70	-6.61*	0.34
Total Problems	33.90	20.94	33.30	18.70	0.59	0.03
<b>DSM-oriented scales</b>						
Affective Problems	3.19	2.66	2.10	2.00	8.99*	0.46
Anxiety Problems	4.33	2.69	3.40	2.50	7.03*	0.36
Pervasive Developmental Problems	3.62	3.27	2.80	2.40	5.55*	0.29
Attention Deficit/Hyperactivity Problems	4.86	3.05	5.00	2.80	-0.94	0.05
Oppositional Defiant Problems	2.70	2.39	3.60	2.50	-7.27*	0.37

\* $p \leq 0.003$ **Table 5** Mean score comparisons and effect sizes for Chinese and U.S. boys and girls on the C-TRF

	Boys				Girls							
	Chinese		U.S.		Chinese		U.S.		t	d'		
	Mean	SD	Mean	SD	Mean	SD	Mean	SD				
<b>Syndromes</b>												
Emotionally Reactive	1.13	1.52	1.5	1.9	-3.60*	0.22	1.13	1.55	1.3	1.9	-1.50	0.10
Anxious/Depressed	1.70	1.86	2.1	2.2	-3.12*	0.20	2.12	2.11	2.2	2.4	-0.55	0.04
Somatic Complaints	0.55	1.03	0.5	1	0.79	0.05	0.57	1.02	0.7	1.2	-1.80	0.12
Withdrawn	2.38	2.71	2.8	3.2	-2.25	0.14	2.15	2.64	2.3	2.9	-0.84	0.05
Attention Problems	4.01	3.18	3.6	3.7	1.89	0.12	2.71	2.85	2.6	3.4	0.54	0.04
Aggressive Behavior	6.40	7.11	6.9	8.5	-1.01	0.06	3.22	4.51	5.3	7.6	-4.97*	0.33
Internalizing	5.73	5.57	6.8	6.6	-2.76	0.17	5.94	5.91	6.4	6.9	-1.10	0.07
Externalizing	10.39	9.52	10.5	11.3	-0.17	0.01	5.93	6.73	8	10.1	-3.63*	0.24
Total Problems	22.31	17.91	23.1	20.9	-0.65	0.04	17.18	15.18	19.6	20.9	-2.02	0.13
<b>DSM-oriented scales</b>												
Affective Problems	1.18	1.66	1.2	2	-0.17	0.01	1.32	1.83	1.2	1.9	1.00	0.06
Anxiety Problems	1.24	1.80	1.1	1.6	1.33	0.08	1.41	1.69	1.3	1.9	0.95	0.06
Pervasive Developmental Problems	3.02	3.09	3.2	3.2	-0.92	0.06	2.79	2.91	2.6	3	1.00	0.06
Attention Deficit/Hyperactivity Problems	5.53	4.39	5.5	5.5	0.10	0.01	3.79	3.78	4.1	4.9	-1.08	0.07
Oppositional Defiant Problems	1.30	1.90	2.1	3	-4.99*	0.32	0.78	1.43	1.7	2.7	-6.35*	0.43

\* $p \leq 0.003$



the U.S. and Chinese samples, the former often scored higher, with small effect sizes (mean=0.29, range=0.20 to 0.43). Specifically, girls in the U.S. sample were found to score significantly higher on Externalizing Problems, Aggressive Behavior, and Oppositional Defiant Problems than Chinese girls. Meanwhile, boys in the U.S. sample scored significantly higher on Emotionally Reactive, Anxious/Depressed, and Oppositional Defiant Problems than Chinese boys.

**Cross-Informant Agreement** Pearson correlations between scale scores of our Chinese CBCL/1.5-5 (parents) and C-TRF (teachers) were positive and significant at  $p < 0.001$ , except for Somatic Complaints. The mean cross-informant correlation was 0.18 (range 0.04–0.25), significantly smaller than the mean of 0.40 (range 0.21–0.58) in the U.S. sample (Achenbach and Rescorla 2000), according to Fisher’s  $z$ -test ( $z = 3.22$ ,  $p < 0.001$ ). Specifically, Chinese parents and teachers showed significantly lower agreement on Emotionally Reactive, Anxious/Depressed, Somatic Complaints, Attention Problems, Aggressive Behavior, Internalizing Problems, Externalizing Problems, Total Problems, Pervasive Developmental Problems, Attention Deficit/Hyperactivity Problems, and Oppositional Defiant Problems (Table 6). Interestingly, Somatic Complaints in the Chinese sample was the only syndrome that yielded a non-significant cross-informant correlation among all correlations in the Chinese and U.S. samples. Furthermore, Externalizing Problems displayed the highest cross-informant correlation in both samples, with 0.25 and 0.58 for the Chinese and U.S. samples respectively.

Paired-sample  $t$ -tests were conducted to examine mean differences between ratings of the various scales in the CBCL/1.5-5 and C-TRF, based upon the common items of the two

questionnaires (82 out of 100 items) for direct comparison (Table 7). Except for Withdrawn, parents reported significantly higher levels of problems than teachers for both boys and girls, mean effect sizes being 0.49 (range=0.19 to 0.72) and 0.50 (range=0.19 to 0.75) respectively. Regarding the two broadband Internalizing and Externalizing Problems, the effect sizes of their differences between parents’ and teachers’ reports were similar across gender (0.49 vs. 0.48 on boys and 0.45 vs. 0.67 on girls respectively).

**Discussion**

CFA on the CBCL/1.5-5 with our Chinese sample confirmed the seven-syndrome and two-higher-order-broadband-problem factor structure derived from Achenbach’s and Rescorla’s (2000) U.S. sample. Likewise, the six-syndrome and two-higher-order-broadband-problem factor structure of the C-TRF was also confirmed in our Chinese sample. Analyses on the CBCL/1.5-5 indicated an independent seventh syndrome, Sleep Problems, which belonged to neither of the two higher-order factors. Given such confirmation, this study, to the best of our knowledge, is the first to report the cross-cultural applicability to a Chinese sample of a taxonomy of syndromal constructs of preschool psychopathology derived in the U.S. from the CBCL/1.5-5 and C-TRF. These findings support their factorial validity and suggest that the two questionnaires’ U.S.-derived taxonomy is appropriate for assessing Chinese preschool children for psychopathology.

Our results suggest that teachers do not seem to rate the two broad-band problems (i.e., internalizing and externalizing problems) as similar, in level or severity, as parents would rate

**Table 6** Comparison of cross-informant agreement (correlation) between the Chinese and American samples (Fisher’s  $z$ -test)

	Chinese <i>r</i>	American <i>r</i>	<i>z</i>	<i>p</i>
<b>Syndromes</b>				
Emotionally Reactive	0.15	0.28	1.82	0.03*
Anxious/Depressed	0.16	0.28	1.68	0.05*
Somatic Complaints	0.04	0.30	3.59	0.00***
Withdrawn	0.23	0.29	0.86	0.20
Attention Problems	0.25	0.51	4.1	0.00***
Aggressive Behavior	0.23	0.55	5.12	0.00***
Internalizing	0.15	0.30	2.11	0.02*
Externalizing	0.25	0.58	5.43	0.00***
Total Problems	0.18	0.50	4.9	0.00***
<b>DSM-Oriented Scales</b>				
Affective Problems	0.12	0.21	1.23	0.11
Anxiety Problems	0.15	0.26	1.53	0.06
Pervasive Developmental Problems	0.25	0.42	2.56	0.01**
Attention Deficit/Hyperactivity Problems	0.18	0.52	5.26	0.00***
Oppositional Defiant Problems	0.17	0.42	3.68	0.00***

All correlations significant at  $p \leq 0.05$ , except for Somatic Complaints in the Chinese sample  
 \* $p \leq 0.05$ , \*\* $p \leq 0.01$ , \*\*\*  $p \leq 0.001$

**Table 7** Mean score comparisons between scales of the CBCL/1.5-5 and C-TRF (based on 82 common items)

	Boys						Girls					
	CBCL Mean	C-TRF Mean	Mean Diff.	SD	t	d'	CBCL Mean	C-TRF Mean	Mean Diff.	SD	t	d'
<b>Syndromes</b>												
Emotionally Reactive	2.12	1.14	0.98	2.27	8.95*	0.43	2.08	1.12	0.96	2.21	8.58*	0.43
Anxious/Depressed	3.32	1.72	1.60	2.68	12.56*	0.60	3.23	2.09	1.14	2.83	7.98*	0.40
Somatic Complaints	1.72	0.55	1.17	2.03	12.06*	0.58	1.76	0.56	1.19	1.96	12.21*	0.61
Withdrawn	2.32	2.00	0.31	2.87	2.30	0.11	2.12	1.86	0.26	2.88	1.78	0.09
Attention Problems	3.01	2.39	0.62	2.36	5.65*	0.26	2.54	1.55	0.99	2.19	9.12*	0.45
Aggressive Behavior	8.58	5.04	3.53	7.06	10.25*	0.50	6.67	2.71	3.96	5.88	13.15*	0.67
Internalizing	9.18	5.35	3.83	7.80	9.61*	0.49	8.92	5.46	3.45	7.74	8.38*	0.45
Externalizing	11.60	7.42	4.18	8.67	9.84*	0.48	9.16	4.26	4.90	7.35	12.90*	0.67
Total Problems	29.81	17.40	12.40	20.92	9.98*	0.59	24.65	13.19	11.46	19.35	9.93*	0.59
<b>DSM-oriented scales</b>												
Affective Problems	1.55	0.96	0.58	1.97	6.32*	0.29	1.44	1.07	0.37	1.97	3.74*	0.19
Anxiety Problems	2.55	1.10	1.44	2.00	14.33*	0.72	2.37	1.22	1.15	2.22	9.75*	0.52
Pervasive Developmental Problems	3.68	2.94	0.74	3.87	3.90*	0.19	3.49	2.71	0.78	3.71	4.18*	0.21
Attention Deficit/Hyperactivity Problems	5.08	2.90	2.18	3.19	14.27*	0.68	4.60	2.16	2.44	3.27	14.85*	0.75
Oppositional Defiant Problems	2.92	1.23	1.70	2.70	13.36*	0.63	2.46	0.74	1.72	2.61	13.32*	0.66

\* $p \leq 0.003$ 

their children. There may be a number of explanations. First, the children's behaviors may simply differ at school and at home. Second, teachers may be more discriminative informants than parents in China and detect differences in the prevalence of internalizing and externalizing problems in the same child. Preschool teachers in China receive compulsory training in preschool psychology (China Education and Research Network 1998–2000) and thus may be more knowledgeable than parents on child psychopathology. Furthermore, given the teachers' more extensive experience with different children, they are likely to acquire greater expertise in differentiating different types of problem behaviors than parents in China, who are only allowed to have one child. Third, the parents may overfocus on their single child, thus making problems seem more visible than they really are. This leads to a higher correlation between internalizing and externalizing problems in the CBCL/1.5-5. Fourth, conversely, teachers may not be more discriminative. Because of the large class size (about 40), they may not have sufficient time to attend to each individual child. They may only be able to identify some problems but miss the others. Since this study has not administered another independent or "gold standard" measure on children's behaviors, we cannot judge the validity of the more correlated CBCL/1.5-5 or less correlated C-TRF. Both may be valid, given that children may have different behaviors at different settings (home and school). Future studies that use objective and independent direct observation other than rating scales may shed some light on this issue.

Cultural differences between the Chinese and U.S. samples were apparent in this study. Before discussing these differences, it should be noted that the age range of our Chinese sample only spans between 4 to 5 years, while the original U.S. sample ranged from 1.5 to 5 years. The different age ranges may attenuate or reduce our cross-cultural differences or similarities. However, in the original U.S. sample (Achenbach and Rescorla 2000), there were no reported significant effects of age on the C-TRF scales, while the effect sizes for age on a small number of the CBCL/1.5-5 scales were mostly in the negligible range of 1 to 2%. These provide at least partial assurance on the comparability of the two samples, despite different age ranges. Second, there may be cross-cultural differences in the interpretation of Likert-type response categories used in the CBCL/1.5-5 and C-TRF. One study, though not involving our present questionnaires, reported wide variations in the interpretation of such categories, even among a demographically homogeneous sample. For example, when college students were asked to define the term "often" in rating their frequency of asking others to read a passage that they wrote, 18% indicated "once or twice per year", 33% "3 to 6 times per year", and 35% "1 to 2 times per month" (Pace and Friedlander 1982). Such differences in interpretation may also exist between U.S. and Chinese raters of the CBCL/1.5-5 and C-TRF. Third, there may be cross-cultural differences in raters' expectations and norms that affect their rating, independently of the behaviors of the

children. Despite the above warnings, our cross-cultural findings should gain more credibility if they are consistent with existing literature or theoretically interpretable.

In the CBCL/1.5-5, significant differences between Chinese and U.S. samples confirm prior findings that Chinese/Asian children experience more internalizing problems whereas Western children experience more externalizing problems (Liu et al. 2001; Weine et al. 1995; Yang et al. 2000; Weisz et al. 1995). The “problem suppression-facilitation model” (Weisz et al. 1987) has been advocated to explain this phenomenon. The model suggests that certain cultural factors suppress the development of specific child problems while facilitating the development of others. It has been speculated that Chinese/Asian children subjected to socialization practices that stress self-control, emotional restraint, submissiveness, and dependency on others’ opinions (e.g. Chao 1995; Chen et al. 1997; Ho and Kang 1984) are more prone to developing internalizing/overcontrolled problems, as opposed to externalizing/undercontrolled problems (Zahn-Waxler et al. 2000; Weisz et al. 1993). Conversely, Western children subjected to socialization practices that encourage assertion of one’s own wills and needs are thought to suffer from a higher risk of developing externalizing/under-controlled problems (Rothbaum et al. 2000; Weisz et al. 2006). Furthermore, the differences reported may be biological. Some evidence has demonstrated ethnic differences between Chinese and American children in their autonomic nervous systems (Kagan et al. 1978), with Chinese children having more inhibited temperaments.

From a cultural context, however, an alternative, opposite view may be that internalizing behaviors are adaptive in Chinese children and may represent “normal” reactions to some dominant cultural forces. In this vein, a higher level of internalizing behaviors does not necessarily reflect the presence of pathology in Chinese children. A similar argument can apply to heightened externalizing behaviors in U.S. children, who live in a culture emphasizing individualistic strivings. Additional research is needed to tease out the relationship between socialization practices and biologically-rooted dispositional characteristics in the development of these two broadband groups of childhood disorders in Eastern and Western cultures.

Cross-cultural differences are also noted in the C-TRF. Once again, we must first note that in this study, the teachers rated all students in the class, about 40 in total. Despite the fact that teacher burden was minimized by spreading ratings throughout the year, it is still unclear whether and how this burden may have affected ratings. Perhaps, scores may be lowered, given that, compared to the CBCL/1.5-5, C-TRF scores were lower on almost all scales. Thus, our C-TRF findings can only partly reproduce those of CBCL/1.5-5, i.e., finding similarly fewer externalizing

problems, but failing to demonstrate more internalizing problems in Chinese preschoolers. However, the above findings are not only found in our study. These cross-cultural differences in teachers’ versus parents’ reports resemble findings from a study with Chinese adolescents (Yang et al. 2000). Previous research has also shown that teachers tend to report a less consistent pattern of cross-cultural differences on Asian children’s behavior problems (Weine et al. 1995; Yang et al. 2000). Yang et al. (2000) argued that Chinese teachers were less reliable than parents in detecting covert internalizing problems because of their relatively less frequent and intense interaction with students.

Ratings of Total Problems by either parents or teachers are mostly similar between Chinese and U.S. samples. For those scales with significant Chinese and U.S. differences, the effect sizes are generally small (see Tables 4 and 5). Weine et al. (1995) found considerable similarities in total problem scores of rural and urban Chinese and U.S. children (ages 6 to 13 years). The authors pointed out that any differences reported were only evident in the TRF and might reflect discrepancies in the school environment, rather than inherent differences between the two cultures as a whole. Similarly, Liu et al. (2001) also found the prevalence of emotional and behavioral problems as measured by the CBCL in Chinese adolescents to be largely consistent with those seen among U.S. adolescents.

This study documents gender differences, with boys scoring significantly higher than girls on all externalizing-related problem scales on both questionnaires, as expected. No gender differences are found on internalizing-related problems on either questionnaire, with the exception of a higher score for girls on Anxious/Depressed in the C-TRF, but having a small effect size of 0.21. Findings correspond well with previous research on developmental psychopathology showing that after the age of 4 years, boys tend to have higher rates of externalizing disorders than girls, while rates of internalizing disorders are similar between the two genders until adolescence (Offord et al. 1987; Angold and Rutter 1992). It is interesting to note that significant gender effects were found on both questionnaires in our Chinese sample, whereas in the U.S. sample, gender effects were only shown on the C-TRF, suggesting a possible interaction effect between country and gender. In another study with older Chinese children, Weine et al. (1995) reported a significant interaction effect between country and gender. Nikapota (2009) suggested in her recent review on cultural issues in child assessment that gender roles, responsibilities, vulnerability, and risk were governed by culture.

The degree of agreement between parents’ and teachers’ ratings of the same emotional and behavioral dimensions in our Chinese sample was relatively low ( $r=0.18$ ) compared to the U.S. sample ( $r=0.40$ ) and to the mean cross-

informant correlation of 0.28 from a major meta-analysis (Achenbach et al. 1987). Nevertheless, our figures are in line with those from previous studies on Chinese children's behavioral problems. Ho et al. (1996) reported in Hong Kong a mean correlation of 0.17 (0.07 to 0.31) between parents' and teachers' ratings on the Rutter's child behavior questionnaires. Deng et al. (2004) reported in Mainland China parent-teacher correlations ranging from 0.12 to 0.36, with the strongest agreement on externalizing and attentional problems and weaker agreement among internalizing behaviors.

The reason why the Chinese in general have a lower cross-informant agreement is not immediately evident. The large class size in both China and Hong Kong, which limits teachers' opportunities to have a thorough knowledge of each individual child, may be one reason. Our Chinese parents reported more problems than teachers on almost all scales across gender and internalizing/externalizing problems. As stated above, the C-TRF ratings may be lowered due to the large number of students rated by the teachers. Further, parents and teachers in Western societies tend to have more frequent contacts, which facilitate information sharing on the children's functioning and thereby render higher agreement between their ratings. For example, parents in Western societies are often encouraged to participate in many aspects of their children's school life, such as helping with school activities and attending parent-teacher meetings. Whether the discrepancy in children's behaviors across situations is genuinely larger among the Chinese than Westerners, or whether there is a larger discrepancy in the knowledge of the children's behaviors across different Chinese informants requires further empirical investigation. A multi-informant approach is now unanimously agreed upon as a standard practice in child assessment (Nikapota 2009), since different informants may provide different but equally valid and complementary perspectives of children's behaviors (Kerr et al. 2007; Verhulst et al. 1994). Despite our lower cross-informant agreement, our findings share similarities with those of the U.S. and other Chinese samples (Deng et al. 2004) that externalizing-related problems have a higher cross-informant agreement than internalizing-related problems. This may be related to the nature of externalizing behaviors being more noticeable.

Interpretation of the present results should take a number of limitations into consideration. First, the taxonomy of problem behaviors identified in this study should not be viewed as a comprehensive syndromal representation of Chinese preschool psychopathology because we have not assessed any possible Chinese-specific problems that are not captured in the two Western-developed questionnaires. Second, results of previous applications of the CBCL and its parallels to school-age children/adolescents in other Chinese communities, such as Hong Kong or Taiwan (e.g., Leung et al. 2006; Yang et al. 2000), may not be necessarily

generalizable to Mainland Chinese preschoolers, given the age differences and the latter two Chinese communities having considerably more exposure to Western influences. Nonetheless, the converging results from studies of different age groups and Chinese communities give support to our current findings. Third, while the Jintan area is in many respects broadly representative of Mainland China, it represents neither a metropolitan city nor the rural countryside. A next step would be to include the preschool children in other geographic regions of Mainland China to increase the generalizability of our findings. Finally, the fact that teachers rate the entire class of students, i.e., multiple children, violates the statistical requirement of independent observations, and this may mitigate results. In addition, since there are approximately 40 students in each class, the teachers may have had difficulty completing the forms for all students. However, the rating of the forms was encouraged to be spread out throughout the semester; so the teachers could have sufficient time to do so.

In summary, our study is the first to include a mixed gender sample of Chinese preschoolers to report on the cross-cultural applicability and validity of the multi-syndrome and two-broadband-problem factor structure of the CBCL/1.5-5 and C-TRF to Chinese. It supports the generalizability of a taxonomy of preschool psychopathology to a cultural and socio-economic group (Chinese) very different from the cultural group (American) from which the constructs are originally derived. We believe that results from the present study can help to establish a multi-cultural test of the taxonomy of preschool psychopathology across many cultures/societies. Lastly, the convergence of the syndrome structure across different societies and the modest country effect on syndrome scores correspond well with recent multicultural findings on older children and adolescents (Ivanova et al. 2007a, b, c; Rescorla Rescorla 2007; Rescorla et al. 2007a, b), rendering support for the universalist's position on child psychopathology.

**Acknowledgements** This study is, in part, supported by NIH/NIEHS K01-ES015 877.

We would like to thank the participating children, their families, and the schoolteachers from Jintan City as well as the Jintan Cohort Study Group.

We would also like to acknowledge Dr. Bu Huang's statistical assistance during the revision process.

## References

- Achenbach, T., & Rescorla, L. (2000). *Manual for the ASEBA preschool forms & profiles*. Burlington: University of Vermont, Research Centre for Children, Youth, & Families.
- Achenbach, T., & Rescorla, L. (2001). *Manual for the ASEBA school-age forms & profiles*. Burlington: University of Vermont, Research Centre for Children, Youth, & Families.
- Achenbach, T., McConaughy, S., & Howell, C. (1987). Child/adolescent behavioral and emotional problems: implications of

- cross-informant correlations for situational specificity. *Psychological Bulletin*, 101, 213–232.
- Achenbach System of Empirically Based Assessment (2009). Translations of psychopathology of preschool children in China 71forms. Accessed March 4, 2009 from <http://www.aseba.org/ordering/translations.html>.
- American Psychiatric Association (APA) (1994). *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*, 4th edn. American Psychiatric Press, Washington.
- Angold, A., & Rutter, M. (1992). Effects of age and pubertal status on depression in a large clinical sample. *Development and Psychopathology*, 4, 5–28.
- Bentler, P. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238–246.
- Browne, N., & 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.
- Canion, G., & Alegria, M. (2008). Psychiatric diagnosis—is it universal or relative to culture? *Journal of Child Psychology and Psychiatry*, 49, 237–250.
- Cederblad, M., Pruksachatkunakorn, P., Boripunkul, T., Intraprasert, S., & Hook, B. (2001). Behavior problems and competence in Thai children and youths: teachers', parents' and subjects' perspectives. *Transcultural Psychiatry*, 38, 64–79.
- Chao, R. (1995). Chinese and European American cultural models of the self reflected in mothers' child rearing beliefs. *Ethos*, 23, 328–354.
- Chen, X., Rubin, K., & Li, B. (1997). Maternal acceptance and social and school adjustment in Chinese children: a four-year longitudinal study. *Merrill-Palmer Quarterly*, 43, 663–681.
- China Education and Research Network (1998–2000). *Teacher education in China (III)*. Accessed June 21, 2009 from <http://www.edu.cn/20010101/21925.shtml>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic.
- Deng, S., Liu, X., & Roosa, M. (2004). Agreement between parent and teacher reports on behavioral problems among Chinese children. *Journal of Developmental and Behavioral Pediatrics*, 25(6), 407–414.
- Egger, H., & Angold, A. (2006). Common emotional and behavioral disorders in preschool children: presentation, nosology, and epidemiology. *Journal of Child Psychology and Psychiatry*, 47(3/4), 313–337.
- Farrington, D. (1989). Early predictors of adolescent aggression and adult violence. *Violence and Victims*, 4, 79–100.
- Ho, D., & Kang, T. (1984). Intergenerational comparisons of child-rearing attitudes and practices in Hong Kong. *Developmental Psychology*, 20, 1004–1016.
- Ho, T., Leung, P., Luk, E., Taylor, E., Bacon-Shone, J., & Mak, F. (1996). Establishing the constructs of childhood behavioral disturbances in a Chinese population: a questionnaire study. *Journal of Abnormal Child Psychology*, 24(4), 417–431.
- Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.
- Ivanova, M., Achenbach, T., Dumenci, L., Rescorla, L., Almqvist, F., Weintraub, S., et al. (2007a). Testing the 8-syndrome structure of the Child Behavior Checklist in 30 societies. *Journal of Clinical Child and Adolescent Psychology*, 36(3), 405–417.
- Ivanova, M., Achenbach, T., Rescorla, L., Dumenci, L., Almqvist, F., Bathiche, M., et al. (2007b). Psychopathology of preschool children in China 77. Testing the teacher's report from syndromes in 20 societies. *School Psychology Review*, 36(3), 468–483.
- Ivanova, M., Achenbach, T., Rescorla, L., Dumenci, L., Almqvist, F., Bilenberg, N., et al. (2007c). The generalizability of the Youth Self-Report syndrome structure in 23 societies. *Journal of Consulting and Clinical Psychology*, 75, 729–738.
- Kagan, J., Kearsley, R., & Zelazo, P. (1978). *Infancy: Its place in human development*. Cambridge: Harvard University Press.
- Kerr, D., Lunkenheimer, E., & Olson, S. (2007). Assessment of child problem behaviors by multiple informants: a longitudinal study from preschool to school entry. *Journal of Child Psychology and Psychiatry*, 48(10), 967–975.
- Leung, P., & Wong, M. (2003). Measures of child and adolescent psychopathology in Asia. *Psychological Assessment*, 15(3), 268–279.
- Leung, P., Kwong, S., Tang, C., Ho, T., Hung, S., Lee, C., et al. (2006). Test-retest reliability and criterion validity of the Chinese version of CBCL, TRF, and YSR. *Journal of Child Psychology and Psychiatry*, 47(9), 970–973.
- Liu, J., & Wuerker, A. (2005). A biosocial bases of violence: implications for nursing research. *International Journal of Nursing Studies*, 42, 229–241.
- Liu, J., McCauley, L., Zhao, Y., Pinto-Martin, J., & Jintan Cohort Study Group. (2010). Cohort profile: The China Jintan Child Cohort Study. *International Journal of Epidemiology*, 39(3), 668–674.
- Liu, X., Sun, Z., Neiderhiser, J., Uchiyama, M., Okawa, M., & Rogan, W. (2001). Behavioral and emotional problems in Chinese adolescents: parent and teacher reports. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(7), 828–836.
- Marsh, H., Hau, K., & 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: A Multidisciplinary Journal*, 11(3), 320–341.
- Moffitt, T. (1993). Adolescence-limited and life-course-persistent antisocial behavior. A developmental taxonomy. *Psychological Review*, 100, 674–701.
- Muthen, L., & Muthen, B. (1998–2007). *Mplus user's guide*. Fifth Edition. Los Angeles: Muthen & Muthen.
- Nikapota, A. (2009). Cultural issues in child assessment. *Child and Adolescent Mental Health*, 14(4), 200–206.
- Offord, D., Boyle, M., Szatmari, P., Rae Grant, N., Links, P., Cadman, D., et al. (1987). Ontario child health study. II. Six-month prevalence of disorder and rates of service utilization. *Archives of General Psychiatry*, 44, 832–836.
- Pace, C. R., & Friedlander, J. (1982). The meaning of response categories: how often is “occasionally”, “often”, and “very often”? *Research in Higher Education*, 17, 267–281.
- Raine, A. (2002). Biosocial studies of antisocial and violent behavior in children and adults: a review. *Journal of Abnormal Child Psychology*, 30, 311–326.
- Rescorla, L. (2007). Assessment of young children using the Achenbach System of Empirically Based Assessment (ASEBA). *Mental Retardation and Developmental Disabilities Research Reviews*, 11, 226–237.
- Rescorla, L., Achenbach, T., Ivanova, M., Dumenci, L., Almqvist, F., Bilenberg, N., et al. (2007a). Behavioral and emotional problems reported by parents of children ages 6 to 16 in 31 societies. *Journal of Emotional and Behavioral Disorders*, 15(3), 130–142.
- Rescorla, L., Achenbach, T., Ginzburg, S., Ivanova, M., Dumenci, L., Almqvist, F., et al. (2007b). Consistency of teacher-reported problems for students in 21 countries. *School Psychology Review*, 36(1), 91–110.
- Rothbaum, F., Pott, M., Azuma, H., Miyake, K., & Weisz, J. (2000). The development of close relationships in Japan and the United

- States: paths of symbiotic harmony and generative tension. *Child Development*, 71(5), 1121–1142.
- Tan, T., Dedrick, R., & Marfo, K. (2007). Factor structure and clinical implications of child behavior checklist/1.5-5 ratings in a sample of girls adopted from China. *Journal of Pediatric Psychology*, 32(7), 807–818.
- Tremblay, R. (2004). Decade of behavior distinguished lecture: development of physical aggression during infancy. *Infant Mental Health Journal*, 25, 399–407.
- Tucker, L., & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1–10.
- Verhulst, F., Koot, H., & Van der Ende, J. (1994). Differential predictive value of parents' and teachers' reports of children's problem behaviors: a longitudinal study. *Journal of Abnormal Child Psychology*, 22(5), 531–546.
- Weine, A., Phillips, J., & Achenbach, T. (1995). Behavioral and emotional problems among Chinese and American children: parent and teacher reports for ages 6 to 13. *Journal of Abnormal Child Psychology*, 23, 619–639.
- Weisz, J., Suwanlert, S., Chaiyasit, W., Weiss, B., Achenbach, T., & Walter, B. (1987). Epidemiology of behavioral and emotional problems among Thai and American children: parents reports for ages 6–11. *Journal of the American Academy of Child and Adolescent Psychiatry*, 26, 890–897.
- Weisz, J., Sigman, M., Weiss, B., & Mosk, J. (1993). Parent reports of behavioral and emotional problems among children in Kenya, Thailand, and the United States. *Child Development*, 64(1), 98–109.
- Weisz, J., Chaiyasit, W., Weiss, B., Eastman, K., & Jackson, E. (1995). A multimedia study of problem behavior among Thai and American children in school: teacher reports versus direct observations. *Child Development*, 66, 402–415.
- Weisz, J., Weiss, B., Suwanlert, S., & Chaiyasit, W. (2006). Culture and youth psychopathology: testing the syndromal sensitivity model in Thai and American adolescents. *Journal of Consulting and Clinical Psychology*, 74(6), 1098–1107.
- Yang, H., Soong, W., Chiang, C., & Chen, W. (2000). Competence and behavioral/emotional problems among Taiwanese adolescents as reported by parents and teachers. *Journal of the American Academy of Child and Adolescent Psychiatry*, 39, 232–239.
- Yu, C., & Muthen, B. (2002). *Evaluation of model fit indices for latent variable and models with categorical an continuous outcomes (Technical Report)*. Los Angeles: UCLA, Graduate School of Education and Information Studies.
- Zahn-Wasler, C., Klimes-Dougan, B., & Slattery, M. (2000). Internalizing problems of childhood and adolescence: prospects, pitfalls, and progress in understanding the development of anxiety and depression. *Development and Psychopathology*, 12, 443–466.