Journal of Abnormal Child Psychology, Vol. 23, No. 5, 1995

Behavioral and Emotional Problems Among Chinese and American Children: Parent and Teacher Reports for Ages 6 to 13

Andrew M. Weine,[1] Jeanne S. Phillips,[2] and Thomas M. Achenbach[1,3]

This study compared behavioral and emotional problems reported by parents and teachers in Chinese urban and rural samples and demographically similar American samples. Parents of 469 6-to-13-year-old children of each nationality completed the Child Behavior Checklist (CBCL). Teachers completed the Teacher's Report Form (TRF). Cross-cultural differences were generally modest in magnitude. Chinese children scored higher on TRF Delinquent Behavior and Anxious/Depressed syndromes, and on Internalizing. American children scored higher on CBCL Aggressive Behavior and TRF Attention Problems syndromes. Boys exhibited more externalizing behaviors across both cultures. The mean correlation between parent and teacher ratings was .36 in the Chinese sample and .29 in the American sample, a nonsignificant difference. Findings indicate considerable similarity between problems reported for children in very different societies.

In the early stages of a field's development, research is often shaped largely by "local" factors. In the study of child psychopathology, such local factors include the conceptual orientation and methodological background of particular workers, the nature of the institutions where they work, funding contingencies, the available subject populations, and the feasibility of various types of research. The methodology and findings emerging from a particu-

Manuscript received in final form August 3, 1994.
This work was supported by NIMH grant MH40305 and by University Associates in Psychiatry, a nonprofit health service and research corporation of the University of Vermont Department of Psychiatry.

The authors are grateful to David Jacobowitz and Catherine Howell for their help in the analyses.

[1]Department of Psychiatry, University of Vermont, Burlington, Vermont 05401.
[2]Department of Psychology, University of Denver, Denver, Colorado 80208.
[3]Address all correspondence to Dr. Achenbach, Department of Psychiatry, University of Vermont, 1 South Prospect St., Burlington, Vermont 05401.

0091-0627/95/1000-0619$07.50/0

lar local context are of unknown generality until tested elsewhere. Although no single approach can take account of all the variations among all local contexts, an advancing field must determine the degree to which methods and findings developed in one context are generalizable to other contexts.

One level of generalization is across the variations that occur within a country. A second level is across societies that differ in language and culture but share similar social structures, values, and views of development and psychopathology. A third level of generalization is across societies that differ more radically in social structure, values, and relevant views. The present study was designed to test the generalizability of a standardized assessment methodology and its findings across societies that differ radically in many ways. This was done by comparing behavioral/emotional problems reported on the same standardized instruments by parents and teachers of children in the People's Republic of China versus the United States.

There has been relatively little psychological research on child behavior in China, probably in large part due to that country's ambivalence over the years toward psychological assessment (Zhang, 1988). Recent studies have reported on various aspects of Chinese children's behavior, such as aggression, gender differences, prevalence of deviant behavior, and the effects of single-child status on behavior (Ekblad, 1989; Matsuura et al., 1993; Tseng et al., 1988; Xin, Chen, Tang, Lin, & McConville, 1992). However, none of these studies provided rigorous comparisons of Chinese children with children in other countries.

Although lacking demographically similar cross-cultural samples with which to make direct statistical comparisons, several studies have made inferential comparisons between children in China and other countries. For example, in a study of 267 nine- to 13-year-old children in Beijing, Ekblad (1989) concluded that aggressive behavior in Chinese children was as stable over time as aggressive behavior in Scandinavian children assessed with the same measure, Olweus' (1975, 1984) Aggression Inventory.

In another study that lacked direct comparisons between demographically similar samples, Matsuura et al. (1993) reported problem prevalence rates for Chinese, Japanese, and Korean children assessed with the Rutter Parent and Teacher Scales (Rutter, Tizard, & Whitmore, 1970). The sample of 2,432 children from Beijing evidenced more antisocial behavior (destructive acts, disobedience, lying, bullying) than neurotic behavior (worried, miserable, fearful). Yet Chinese parents reported less deviant behavior overall than did Korean and Japanese parents. Chinese parents also reported less deviant behavior than parents in other prevalence studies using the Rutter scales in Britain (Rutter et al., 1970) and New Zealand (McGee, Silva, & Williams, 1984). When rated by teachers, Chinese and Japanese children showed lower rates of deviant behavior than children studied in

Britain (Rutter et al., 1970), Uganda (Minde, 1975), and New Zealand (Matsuura et al., 1993; McGee et al., 1984).

The Child Behavior Checklist (CBCL; Achenbach, 1991b) and the Teacher's Report Form (TRF; Achenbach, 1991c) have been used to obtain standardized reports of children's behavior problems and competencies in several countries, including China. One large study involved administering CBCLs to parents of 3,000 four- and 5-year-olds in the Shanghai area (Xin et al., 1992). Boys were reported to manifest more delinquent behavior and hyperactivity than girls, while girls were reported to manifest more somatic complaints, anxious behavior, and depressed behavior. Commenting on cross-cultural implications, the authors observed that the 98th percentile cutoff scores for Shanghai children were similar to those for U.S. children. However, the lack of direct statistical comparisons with demographically similar American children limits cross-cultural conclusions. This study also lacked teacher reports, an important source of information that may contrast with parent reports.

Single-child status has been of interest in several studies. Using the CBCL, the Shanghai study (Xin et al., 1992) detected no distinctive psychopathology among children who had no siblings. However, in a study of 697 urban and rural preschool children in the Shanghai region, Tseng et al. (1988) found that girls who had no siblings obtained slightly higher scores on factors derived from the Chinese CBCL data concerning depressive, moody, and temperamental behavior. Furthermore, this study found significant interactions between urban versus rural status and single-child status. Rural girls with no siblings had higher scores on a "temper" factor than those who lived in the city. By contrast, rural girls with siblings had lower "temper" scores than those who lived in the city.

Previous studies have made direct statistical comparisons of CBCL scores obtained by American children with those obtained by children from Thailand, Puerto Rico, Australia, France, Greece, Jamaica, Kenya, and the Netherlands. The strongest cross-cultural similarity was found between 3,333 American and Dutch 4- to 16-year-olds, for whom the mean Total Problem scores differed by less than half a point on a 240-point scale (Achenbach, Verhulst, Baron, & Akkerhuis, 1987b). In a comparison between 960 Thai and American 6- to 11-year-olds, Thai children obtained significantly higher mean Total Problem scores than U.S. children (24.2 vs. 20.8), primarily reflecting higher Thai scores on internalizing items (Weisz et al., 1987). However, this difference of 3.4 points in the Total Problem score accounted for only 1.28% of variance and was a small effect according to Cohen's (1988) criteria. Furthermore, a more recent national sample of American children obtained mean CBCL problem scores quite similar to those obtained by the Thai children (Achenbach, 1991b).

Comparisons between 1,448 Puerto Rican and mainland U.S. 4- to 16-year-olds showed significantly higher mean CBCL Total Problem scores in Puerto Rico than in the U.S. (34.9 vs. 20.0), a medium effect accounting for 12.9% of the variance (Achenbach, Bird et al., 1990). In a comparison between 2,600 Australian and American 4- to 16-year-olds, Australian children obtained higher CBCL Total Problem scores than American children (31.6 versus 20.1), a medium effect accounting for 11% of the variance (Achenbach, Hensley, Phares, & Grayson, 1990). In a comparison between parent reports on 1,842 matched French and American children, American children tended to receive higher competence scores, while French children tended to receive higher problem scores (Stanger, Fombonne, & Achenbach, 1994). Scores were higher for French than American children on more internalizing items than externalizing items, though nearly all nationality differences were small. In a comparison between parent reports of 356 Greek and American children, American children again tended to receive higher competence scores, while Greek children tended to receive higher problem scores (MacDonald, Tsiantis, Achenbach, Motti-Stefanidi, & Richardson, 1995). Higher problem scores in the Greek sample were partly due to the tendency of Greek parents to use extreme item scores.

A study comparing CBCLs for 360 Jamaican and 946 U.S. children found few cross-cultural differences (Lambert, Knight, Taylor, & Achenbach, 1994). In rural samples of Embu children in Kenya, Thai children, African-American children, and Caucasian-American children, Embu and Caucasian-American children had significantly higher CBCL Total Problem scores than Thai children (Weisz, Sigman, Weiss, & Mosk, 1993). Cross-cultural comparisons of teachers' ratings have yielded very similar TRF scores in the Netherlands and the U.S. (Achenbach, Verhulst, Edelbrock, Baron, & Akkerhuis, 1987c), but significantly higher TRF scores in Puerto Rico (Achenbach, Bird et al., 1990) and Thailand (Weisz et al., 1989) than in the U.S.

The present study compared behavioral and emotional problems assessed with the same standardized methods in a Chinese sample and a demographically similar American sample. Parents of children in both countries completed the CBCL and their teachers completed the TRF. Used together, these measures enabled us to compare children's functioning not only in very different societies but also in different contexts in each society, as reported by adults who play very different roles in relation to the children. In view of the interactions found between urbanicity and children's problems in China (Tseng et al., 1988), we also tested differences between urban and rural samples with respect to both parent and teacher ratings. By comparing groups of demographically similar Chinese and American children assessed by both parents and teachers, the current study

tested the extent to which assessment methods developed in one society would produce similar or different results in a radically different society.

METHOD

Instruments

CBCL. The CBCL (Achenbach, 1991b) obtains parents' reports on 20 competence items, 118 specific behavioral/emotional problem items, and 2 open-ended problem items. It is scored for three competence scales (Activities, Social, and School), total competence, eight cross-informant syndromes, Internalizing and Externalizing groupings of syndromes, and a total problem score. The eight cross-informant syndromes were empirically derived from parent, teacher, and self-ratings (Achenbach, 1991a). These syndromes are designated as Aggressive Behavior, Anxious/Depressed, Attention Problems, Delinquent Behavior, Social Problems, Somatic Complaints, Thought Problems, and Withdrawn. Extensive evidence has been presented for the reliability and validity of the CBCL (Achenbach, 1991b; Brown & Achenbach, 1995).

TRF. The TRF (Achenbach, 1991c) is the teacher-report companion measure to the CBCL. It includes items for rating academic performance, four adaptive characteristics, 118 specific behavioral/emotional problems, and 2 open-ended items for additional problems. The problem items are scored like those on the CBCL, but employ a 2-month baseline in contrast to the 6-month baseline specified for rating the CBCL items. Ninety-three TRF items have counterparts on the CBCL, although the wording differs slightly, such as referring to "pupils" instead of "children." Twenty-five CBCL items are replaced on the TRF with items that are more appropriate for teachers, as detailed by Achenbach (1991c).

American Sample. American subjects were drawn from a national longitudinal sample assessed in the spring of 1989. Details of the procedures for obtaining the data have been provided by McConaughy, Stanger, and Achenbach (1992). The original sample was first assessed in 1986, when it comprised approximately 100 children of each sex at each age from 4 to 16 years selected to be representative of the U.S. population with respect to socioeconomic status (SES), ethnicity, geographic region, and urban, suburban, or rural residence. Subjects were excluded if they were mentally retarded, or physically handicapped, or if no English-speaking parent or parent-surrogate was available for the interview. In 1986, the overall completion rate for interviews sought on eligible children was 92.1% (Achenbach, Howell, Quay, & Conners, 1991, provide details of the 1986 survey).

In 1989, when subjects were 7 to 19 years old, the CBCL was administered in a home interview to a parent or parent-surrogate of each subject. Interviews were completed for 90.2% of the eligible subjects, for an overall rate of 83.1% for completed assessments of subjects who had been targeted in the 1986 sample (i.e., 92.1% in 1986 × 90.2% in 1989 = 83.1%). In addition, with parental permission, the TRF was completed by teachers of children who attended school. In the 1989 follow-up, the TRF was completed for 75.8% of the eligible subjects 7 years of age or older.

Data for 4- to 6-year-olds were obtained by identifying households in the 1989 assessment that were occupied by children in this age range, in addition to the original subjects who were now 7 to 19 years old. If there was more than one eligible 4- to 6-year-old in the household, the one whose next birthday was closest to the interview date was selected. The CBCL was administered to a parent or parent-surrogate in the same manner as for the older subjects. CBCLs were obtained for 398 (94.8%) of the 420 eligible 4- to 6-year-olds, and TRFs were obtained for 148 (79.6%) of the 186 four- to 6-year-olds who were attending school.

For the present study, candidate subjects included the 998 six- to 13-year-olds with CBCLs and TRFs who had missing data on less than nine problem items on both forms. These subjects included some who had received mental health services within the past 12 months. Although our other cross-cultural comparisons have excluded such children, the lack of child mental health services for the Chinese sample argued for retaining referred American children in the analyses.

Chinese Sample. Chinese translations of the CBCL and TRF were checked and corrected via back-translation and pilot testing with Chinese respondents to communicate the original meaning as accurately but idiomatically as possible. The Chinese data were collected in 1989-1990 in the Fujian province of the south coastal region. Five of the six schools selected were in an urban setting of over one million people. Of these five schools, two were considered to be very selective and of high quality, three were considered to be ordinary and of varying quality. The sixth school served a rural and small-town area.

The target sample included 6- to 13-year-old Chinese children whose teachers and parents were willing to provide information on their behavior. For parents who were illiterate or semiliterate (5 in the urban sample and 75 in the rural sample), the CBCL was read aloud by the teacher or principal to the parent. Other parents completed CBCLs at home and returned them in unmarked sealed envelopes. For their participation, teachers were paid and parents were rewarded with gifts to their children's classes. In half the schools, it was possible to obtain data from children in each grade level from first to sixth grade. In the remaining schools, various grade levels

were sampled; when possible, samples were obtained from every other grade. Teachers were asked to randomly sample by gender and to obtain total samples of 20 in their classes. However, a few classes provided data from more than 20 children.

Research Design

Matched Chinese and American Samples. Urban and rural samples were analyzed separately. The American urban sample came from incorporated urban areas with populations ranging up to over one million, while the rural sample came from unincorporated rural areas plus incorporated areas having populations under 2,500. Chinese children were classified according to the urban versus rural locations of their schools.

Separately for urban and rural groups, American children were matched to Chinese children by gender, age, and, to the degree possible, SES. For the urban samples in both countries, there were roughly equal numbers of boys and girls, while the rural samples in both countries were 56% male and 44% female.

Lacking any well-established system for coding occupational categories in China, we used the educational attainment of the better educated parent to estimate SES. There were 17 levels of education, ranging from the parent being illiterate or semiliterate, to completing various levels of primary, middle, or vocational school, college, or postgraduate studies, which we divided into three groups: lowest SES ranged from illiteracy to some years of the Chinese equivalent of senior high school; middle SES ranged from high school graduation to some years of the Chinese equivalent of 4-year liberal arts college or similar higher education; highest SES ranged from 4-year college graduation to completion of postgraduate studies.

For the American sample, SES was scored on Hollingshead's 9-step scale for the occupation of the parent holding the higher status job, if both had paid employment (Hollingshead, 1975). The occupational scores were grouped as follows: Lowest SES occupational scores ranged from 10 to 40; middle SES scores ranged from 41 to 70; highest SES ranged from 71 to 90. Two digits were used to code occupational status, because occupations that were not clearly codable on Hollingshead's scale were given the mean of the most likely codes (with decimal points omitted).

Of 500 Chinese children having both CBCL and TRF data, 364 urban and 105 rural children were precisely matched to American children for gender and age. Of these children, 84% were matched precisely to American children for SES, 12% were matched within one SES level, and 4% within two SES levels.

RESULTS

For analysis, the children were divided into age groups 6 to 7, 8 to 9, 10 to 11, and 12 to 13. For each sample (urban and rural) and each measure (CBCL and TRF), we performed 2 × 2 × 3 × 4 (Country × Gender × SES Group × Age Group) analyses of variance (ANOVAs) on raw scores for the eight syndromes, internalizing, externalizing, and total problems. Type III sums of squares were used for main and interaction effects, whereby each significant effect was tested after controlling for all other effects.

In addition to evaluating statistical significance at $p < .05$, we evaluated the magnitude of effect sizes according to Cohen's (1988) criteria for ANOVA. Effects accounting for 1 to 5.9% of the variance are considered small, 5.9 to 13.8% medium, and >13.8% large. Because some effects may be significant by chance, we have indicated with superscripts the smallest significant effects in each set of comparisons up to the proportion expected by chance, using a $p < .05$ protection level (Sakoda, Cohen, & Beall, 1954). The smallest effects are the ones that are most likely to be significant by chance.

Nationality Differences

Tables I and III show the results of the ANOVAs on syndrome scores, Internalizing and Externalizing, and Total Problems, for the urban and rural samples, respectively. Significant nationality differences are shown in the first column of these tables. Tables II and IV show the means and standard deviations for the urban and rural samples, listed for boys and girls in each country.

Urban Samples. For the urban samples, there was no significant difference on CBCL Total Problems between Chinese and American samples (Chinese $M = 27.59$, $SD = 16.43$; American $M = 28.11$, $SD = 17.53$). Chinese CBCL scores were significantly higher than American scores only on the Somatic Complaints syndrome, whereas U.S. scores were significantly higher than Chinese scores on the Anxious/Depressed and Aggressive Behavior syndromes, plus the Externalizing grouping. All of the significant effects were small, accounting for 2 to 4% of the variance. On the TRF, Chinese scores were significantly higher than American scores for five syndromes, Internalizing, Externalizing, and Total Problems (Total Problems: Chinese $M = 29.41$, $SD = 22.70$; American $M = 24.47$, $SD = 24.81$). U.S. scores were significantly higher than Chinese on the Attention Problems syndrome. All significant nationality effects on the TRF scales were small, accounting for ≤3% of the variance.

Table I. Percent of Variance Accounted for by Significant ($p < .05$) Effects on Scale Scores for Matched Chinese ($n = 364$) and American ($n = 364$) Urban Samples

Scale	Nationality[a]	Gender[b]	Age[c]	Country × Gender[d]	Country × SES[e]
Child Behavior Checklist					
Withdrawn	—	—	$2^{C,f}$	—	—
Somatic Complaints	$2^{C,f}$	1^{G}	—	1^{CG}	—
Anxious/Depressed	4^{A}	—	—	—	—
Thought Problems	—	—	—	—	—
Social Problems	—	—	$1^{MY,f}$	$<1^{CB,f}$	—
Attention Problems	—	2^{B}	—	—	—
Delinquent Behavior	—	1^{B}	—	$<1^{CB,f}$	—
Aggressive Behavior	3^{A}	$<1^{B,f}$	—	$<1^{AG}$	—
Internalizing	—	—	—	$<1^{AG}$	—
Externalizing	$2^{A,f}$	$1^{B,f}$	—	$<1^{AB}$	—
Total Problems	—	—	—	$<1^{CB}$	—
Teacher's Report Form					
Withdrawn	2^{C}	—	—	—	$1^{CH,f}$
Somatic Complaints	—	—	—	—	—
Anxious/Depressed	2^{C}	—	2^{MO}	—	—
Thought Problems	$<1^{C}$	—	$1^{MY,f}$	$<1^{CB}$	1^{CM}
Social Problems	3^{C}	$<1^{B,f}$	1^{MY}	—	—
Attention Problems	$<1^{A,f}$	2^{B}	—	—	—
Delinquent Behavior	2^{C}	2^{B}	—	2^{CB}	$<1^{CH}$
Aggressive Behavior	—	2^{B}	—	$1^{CB,f}$	—
Internalizing	2^{C}	—	2^{MO}	—	$1^{CH,f}$
Externalizing	$<1^{C,f}$	2^{B}	—	1^{CB}	—
Total Problems	$<1^{C}$	$2^{B,f}$	$1^{MY,f}$	$<1^{CB,f}$	1^{CM}

[a]C = Chinese scored higher; A = Americans scored higher.
[b]B = boys scored higher; G = girls scored higher.
[c]MY = middle youngest age group (8 to 9) scored higher; MO = middle oldest age group (10 to 11) scored higher. The two remaining age groups are the youngest (6 to 7) and the oldest (12 to 13).
[d]CB = Chinese boys scored higher; CG = Chinese girls scored higher; AB = American boys scored higher; AG = American girls scored higher.
[e]CH = Chinese high-socioeconomic status (SES) group scored higher. CM = Chinese middle-SES group scored higher.
[f]Not significant when corrected for number of analyses.

Rural Samples. For the rural samples, there was no significant difference between Chinese and American samples on CBCL Total Problems (Chinese $M = 31.28$, $SD = 14.53$; American $M = 26.01$, $SD = 19.78$). However, Chinese CBCL scores were significantly higher than American scores on the Withdrawn and Delinquent Behavior syndromes, with effects of 3% and 7%, respectively. American CBCL scores were significantly higher than Chinese scores on the Somatic Complaints and Aggressive Be-

Table II. Scale Scores for Chinese and American Urban Samples (*SD* in Parentheses)[a]

	CBCL				TRF			
	China		U.S.		China		U.S.	
Scale	Male (n = 184)	Female (n = 180)	Male (n = 184)	Female (n = 180)	Male (n = 184)	Female (n = 180)	Male (n = 184)	Female (n = 180)
Withdrawn	2.45	2.54	2.06	2.46	2.68	3.38	1.99	2.29
	(2.17)	(2.17)	(1.89)	(2.08)	(3.07)	(3.18)	(3.06)	(3.04)
Somatic Complaints	1.87	2.07	0.80	1.69	0.74	0.48	0.47	0.67
	(1.89)	(1.94)	(1.23)	(2.15)	(1.56)	(0.94)	(1.03)	(1.63)
Anxious/Depressed	2.94	2.76	3.85	4.56	4.91	6.19	3.73	4.11
	(2.96)	(2.87)	(3.24)	(3.78)	(4.39)	(4.42)	(4.39)	(4.95)
Thought Problems	0.74	0.48	0.66	0.71	0.78	0.48	0.38	0.47
	(1.13)	(0.87)	(1.03)	(1.24)	(1.37)	(0.98)	(0.99)	(1.05)
Social Problems	2.83	2.47	2.08	2.56	4.08	3.12	2.24	2.18
	(2.15)	(1.78)	(1.97)	(2.09)	(3.60)	(2.98)	(3.13)	(3.27)
Attention Problems	4.29	3.33	3.78	3.09	8.80	5.30	9.53	6.78
	(3.07)	(2.71)	(3.11)	(3.09)	(7.13)	(5.83)	(8.92)	(7.75)
Delinquent Behavior	2.57	1.62	1.86	1.66	2.90	1.29	1.46	1.11
	(2.14)	(1.43)	(2.03)	(1.79)	(2.95)	(1.58)	(2.29)	(1.97)
Aggressive Behavior	7.70	5.71	8.68	8.86	9.55	4.19	6.68	5.07
	(5.21)	(4.34)	(5.65)	(6.42)	(10.15)	(5.31)	(9.04)	(8.35)
Internalizing	7.21	7.30	6.55	8.54	8.18	9.83	5.94	6.83
	(5.34)	(5.48)	(4.77)	(6.31)	(7.20)	(6.54)	(6.62)	(7.85)
Externalizing	10.27	7.33	10.54	10.52	12.45	5.48	8.14	6.17
	(6.72)	(5.24)	(7.18)	(7.73)	(12.74)	(6.28)	(10.94)	(9.86)
Total Problems	29.69	25.43	26.67	29.57	34.51	24.21	26.39	22.51
	(17.19)	(15.37)	(15.79)	(19.09)	(25.74)	(17.72)	(25.11)	(24.42)

[a]CBCL = Child Behavior Checklist; TRF = Teacher's Report Form.

havior syndromes, with effects ≤3%. On the TRF, there was no significant difference on Total Problems between Chinese and American rural samples (Chinese M = 38.98, SD = 20.19; American M = 29.39, SD = 29.70). Chinese scores were significantly higher on the Anxious/Depressed and Delinquent Behavior syndromes and the Internalizing grouping, whereas American scores were higher on the Somatic Complaints and Attention Problems syndromes. The effect for Delinquent Behavior was medium, accounting for 7% of the variance, while all other effects were small, accounting for ≤5% of the variance.

Gender Differences

Urban Samples. As Table I shows, urban boys obtained significantly higher CBCL scores on the Attention Problems, Delinquent Behavior, and Aggressive Behavior syndromes, and on Externalizing. Girls scored significantly higher on the CBCL Somatic Complaints syndrome, but the effects ranged only up to 2%. On the TRF, boys scored significantly higher than girls on four of the 11 syndromes, Externalizing, and Total Problems, accounting for ≤2% of the variance.

Rural Samples. As Table III shows, rural girls scored significantly higher on the CBCL Somatic Complaints and Anxious/Depressed syndromes and on Internalizing, accounting for 2 to 4% of the variance. On the TRF, boys obtained significantly higher scores on the Attention Problems, Delinquent Behavior, and Aggressive Behavior syndromes, Externalizing, and Total Problems, accounting for ≤5% of the variance.

Age Differences

Urban Samples. Among urban children, the number of significant effects for age on the CBCL did not exceed chance expectations. On the TRF, the 10- to 11-year-old group scored highest on the Anxious/Depressed syndrome and Internalizing grouping, while the 8- to 9-year-old group scored highest on the Thought Problems and Social Problems syndromes, and Total Problems, with effects ranging up to 2%.

Rural Samples. Among rural children, the oldest group (12 to 13 years old) scored highest on the CBCL Anxious/Depressed syndrome and Internalizing. The 8- to 9-year-old group scored highest on the Social Problems syndrome and Total Problems. The youngest group (6 to 7 years old) scored highest on the Attention Problems syndrome. On the TRF, there were fewer significant effects of age than expected by chance.

Table III. Percent of Variance Accounted for by Significant ($p < .05$) Effects on Scale Scores for Matched Chinese ($n = 105$) and American ($n = 105$) Rural Samples

Scale	Nationality[a]	Gender[b]	Age[c]	Country × Gender[d]	Gender × Age[e]
Child Behavior Checklist					
Withdrawn	$3^{C,g}$	—	—	$2^{CG,g}$	—
Somatic Complaints	4^{A}	$2^{G,g}$	—	—	—
Anxious/Depressed	—	4^{G}	5^{O}	3^{CG}	$4^{MYG,g}$
Thought Problems	—	—	—	—	—
Social Problems	—	—	$4^{MY,g}$	—	—
Attention Problems	—	—	6^{Y}	—	—
Delinquent Behavior	7^{C}	—	—	2^{CG}	—
Aggressive Behavior	$2^{A,g}$	—	—	—	6^{MOB}
Internalizing	—	$4^{G,g}$	4^{O}	$2^{CG,g}$	$4^{MYG,g}$
Externalizing	—	—	—	—	5^{MOB}
Total Problems	—	—	$4^{MY,g}$	—	5^{OB}
		Country × Age			
Teacher's Report Form					
Withdrawn	—	—	—	$6^{AY,g}$	
Somatic Complaints	$2^{A,g}$	—	—	6^{AY}	
Anxious/Depressed	5^{C}	—	—	$6^{CMO,g}$	
Thought Problems	—	—	—	—	
Social Problems	—	—	—	—	
Attention Problems	3^{A}	5^{B}	—	—	
Delinquent Behavior	7^{C}	$3^{B,g}$	—	—	
Aggressive Behavior	—	4^{B}	—	—	
Internalizing	$2^{C,g}$	—	—	8^{CO}	
Externalizing	—	4^{B}	—	—	
Total Problems	—	$3^{B,g}$	—	—	

[a]C = Chinese scored higher; A = Americans scored higher.
[b]B = boys scored higher; G = girls scored higher.
[c]O = oldest age group (12 to 13) scored higher; MY = middle youngest age group (8 to 9) scored higher; youngest age group (6 to 7) scored higher.
[d]CB = Chinese boys scored higher; CG = Chinese girls scored higher; AB = American boys scored higher; AG = American girls scored higher.
[e]OB = oldest (12 to 13) boys scored higher; MO = middle oldest (10 to 11) boys scored higher; MYG = middle youngest (8 to 9) girls scored higher.
[f]CO = Chinese oldest age group (12 to 13) scored higher; CMO = Chinese middle youngest age group (8 to 9) scored higher; AY = American youngest age group scored higher (6 to 7).
[g]Not significant when corrected for number of analyses.

SES Differences

For all samples, there were fewer main effects of SES than expected by chance.

Interactions

Tables I and II list interactions that exceeded chance expectations. Thus, for instance, Country × Gender interactions were the only ones that exceeded chance expectations in the CBCL analyses of urban children. Overall, only one significant interaction was consistent across either urban and rural samples or parent and teacher reports. In parent and teacher reports for the urban samples, Chinese boys scored higher than both Chinese girls and American boys and girls on the Delinquent Behavior syndrome and Total Problems.

Comparison of Problem Items

Nationality Differences. To assess cross-cultural differences in further detail, we did ANOVAs on every problem item like those we had done on the scale scores. Results were examined for effects meeting Cohen's (1988) criteria for medium effect sizes (≥5.9% of variance) that were evident in three out of four analyses (urban CBCL and TRF; rural CBCL and TRF). Only one item met this criterion: 64—*Prefers being with younger children.* In all four analyses, Chinese children scored significantly higher than American children, with effect sizes ranging from 6 to 8%. There were no other significant cross-cultural differences of this size or consistency.

Relations Between Parent and Teacher Ratings

To test the relations between parent and teacher reports, we computed Pearson *r* between parent and teacher ratings of the Chinese and American children on each scale. The *r*-values between Chinese parents and teachers ranged from .05 to .56, while the *r*-values between American parents and teachers ranged from .03 to .46. According to Fisher's *z*-test for differences between *r*-values, Chinese parents and teachers showed significantly better agreement on the Somatic Complaints syndrome ($r = .28$ vs. .09), the Delinquent Behavior syndrome ($r = .56$ vs. .40), and Externalizing ($r = .47$ vs. .37), just one more than the two significant differences expected by chance among 11 analyses. The *r*-values between parents and teachers were also nonsignificantly higher in the Chinese than the American sample on all other scales, except Social Problems. The mean *r* computed by *z*-transformation across all 11 scales was .36 in the Chinese sample and .29 in the American sample, a nonsignificant difference.

To further compare the consistency between parent and teacher reports between countries, we performed 2 × 2 (Country × Parent vs.

Table IV. Scale Scores for Chinese and American Rural Samples (*SD* in Parentheses)[a]

	CBCL				TRF			
	China		U.S.		China		U.S.	
Scale	Male (n = 59)	Female (n = 46)	Male (n = 59)	Female (n = 46)	Male (n = 59)	Female (n = 46)	Male (n = 59)	Female (n = 46)
Withdrawn	2.44	3.74	1.97	1.78	2.41	4.00	2.88	2.26
	(1.88)	(2.77)	(2.04)	(2.01)	(2.51)	(2.76)	(3.17)	(3.12)
Somatic Complaints	0.64	0.87	1.20	1.52	0.39	0.74	0.93	0.98
	(0.85)	(0.93)	(1.58)	(2.01)	(0.77)	(1.08)	(2.31)	(2.06)
Anxious/Depressed	3.17	5.22	3.88	3.65	5.86	8.30	3.47	4.04
	(2.61)	(3.79)	(4.69)	(3.90)	(4.88)	(5.28)	(4.07)	(5.07)
Thought Problems	0.51	0.35	0.80	0.52	0.17	0.33	0.68	0.41
	(0.84)	(0.60)	(1.21)	(1.13)	(0.42)	(0.70)	(1.15)	(0.83)
Social Problems	3.03	3.07	2.17	1.72	3.90	4.35	2.88	2.52
	(2.32)	(2.19)	(2.22)	(1.92)	(3.13)	(3.19)	(4.01)	(3.26)
Attention Problems	3.97	2.74	4.22	2.98	9.64	6.83	11.47	7.93
	(2.67)	(2.38)	(3.34)	(3.22)	(6.35)	(6.02)	(9.96)	(8.52)
Delinquent Behavior	5.61	2.89	1.86	1.37	5.58	3.11	2.19	1.02
	(3.52)	(2.22)	(1.80)	(1.92)	(3.85)	(2.44)	(2.99)	(1.61)
Aggressive Behavior	8.56	5.67	8.86	7.50	10.12	5.61	9.19	4.59
	(6.53)	(4.37)	(5.58)	(6.49)	(7.80)	(4.71)	(12.04)	(6.95)
Internalizing	6.14	9.63	6.85	6.83	8.56	12.93	6.95	7.07
	(3.68)	(5.97)	(6.69)	(6.58)	(6.48)	(7.15)	(7.65)	(8.53)
Externalizing	14.17	8.57	10.73	8.87	15.69	8.72	11.37	5.61
	(9.07)	(5.84)	(6.74)	(7.92)	(11.21)	(6.70)	(14.91)	(7.97)
Total Problems	33.44	28.50	27.44	24.17	41.49	35.76	34.17	23.26
	(14.71)	(13.95)	(18.89)	(20.93)	(20.59)	(19.40)	(33.51)	(22.88)

[a]CBCL = Child Behavior Checklist; TRF = Teacher's Report Form.

Teacher) ANOVAs on Total Problems separately for the urban and rural samples. There was a significant Country × Rater interaction for the urban sample ($p < .05$), but not for the rural sample. Chinese teachers rated urban children higher on Total Problems than did Chinese parents or American teachers or parents, but it was a very small effect, accounting for <1% of variance.

DISCUSSION

Cross-Cultural Comparisons

For both the urban and rural samples, differences between cultures were generally modest. For Total Problems, there were no significant differences between American and Chinese CBCL scores in either the urban or rural sample, nor between TRF scores in the rural samples. The only significant difference was between TRF scores for the urban sample, but this accounted for <1% of the variance. To compare the proportion of Chinese and American subjects who scored in the clinical and borderline ranges for Total Problems (T-score ≥ 60), we also performed 2 × 2 Chi Square analyses (Country × Proportion scoring <60 vs. ≥60 separately for urban CBCL and TRF, and rural CBCL and TRF). Three of the four showed no significant differences. The fourth (rural CBCL) was nominally significant at $p = .017$, but the Bonferroni correction increased this to $p = .068$. Categorical comparisons thus showed considerable similarities between American and Chinese CBCL and TRF scores. We also computed correlations between mean scores obtained for each country on all problem items of the CBCL and TRF. Pearson r-values between countries on all problem items were as follows: for the urban CBCL sample, $r = .66$; for rural CBCL, $r = .44$; for urban TRF, $r = .72$; and for rural TRF, $r = .44$. Fisher's z-test for differences between r values showed that the cross-cultural similarity between the mean item scores was significantly greater for the urban samples than the rural samples on both the CBCL and TRF (both $p < .05$).

There was no significant difference between the mean of all parent–teacher correlations for the Chinese versus American samples, $r = .36$ versus .29. Because Chinese teachers and principals administered the CBCL to the 80 parents with poor reading skills, teachers' knowledge of these parents' responses might have contributed to the slightly higher parent–teacher correlations in China than the U.S. However, the parent–teacher correlations in both countries were very similar to the mean $r = .27$ found in meta-analyses of many previous studies (Achenbach, McConaughy, & Howell, 1987a).

Syndrome Scales. Urban and rural Chinese children obtained significantly higher scores than American children on the TRF Delinquent and Anxious/Depressed syndromes and on Internalizing, with the differences accounting for small to medium percentages of variance. In both urban and rural samples, American children were rated significantly higher on the Aggressive Behavior syndrome by their parents than were Chinese children, though the effects were small. American children in both urban and rural samples also obtained significantly higher scores on the TRF Attention Problems syndrome, though these effects were small as well.

Considering the overall similarity in total problem scores, the cross-cultural differences emerging in a small proportion of the syndrome comparisons suggest only minor differences in the patterns of problems for Chinese versus American children. The findings that were most consistent suggest somewhat more overt behavior problems among American children, as reflected in the CBCL Aggressive syndrome and TRF Attention Problems syndrome scores. By contrast, the problems on which Chinese children scored higher were more covert and were confined to school, as reflected in the TRF Anxious/Depressed, Internalizing, and Delinquent scores.

Weisz et al. (1987) noted that some apparent cross-cultural differences may reflect differences in thresholds for reporting particular kinds of problems. Yet even the most consistent differences between Chinese and American scores were not consistent across parents and teachers. This suggests that the differences do not merely reflect differences between Chinese versus American thresholds for reporting particular kinds of problems. Instead, the findings are apt to encompass both within- and between-culture variations in children's functioning, plus possible variations in the perceptions of different kinds of informants. Thus, for example, Chinese school settings may evoke more covert problem behavior, Chinese teachers may be more sensitive to such behavior, or both may contribute to the higher Chinese scores on three TRF scales. American schools, by contrast, may evoke more attention problems, American teachers may be more sensitive to such problems, or both may contribute to the higher American scores on the TRF Attention Problems scale. Although we cannot separate these possible contributions, the overall picture is one of remarkably small differences in reported problems, despite the many factors that could cause differences.

The current results are consistent with results from comparisons of Thai and U.S. children (Weisz et al., 1989, 1987), which provide the closest comparison to our study, both in terms of the proximity and similarity of the populations and use of the same measures. In the study of Thai teacher reports (Weisz et al., 1989), Thai children were rated by their teachers as being higher than U.S. children on both Internalizing and Externalizing.

Similarly, in the current study, Chinese children were rated higher by their teachers than American children on the Anxious/Depressed syndrome and Internalizing, but also on the Delinquent Behavior syndrome of the Externalizing grouping.

A further similarity between the Thai and Chinese studies concerns some of the specific differences between parent and teacher reports. In both comparisons, Asian teachers (whether Thai or Chinese) generally rated children higher on Total Problems than Asian parents or American teachers or parents (Weisz et al., 1989). For the urban sample in this study, Chinese youngsters were about the same as Americans when parent ratings were used (Chinese M = 27.6, U.S. M = 28.1) but 5 points higher than American youngsters when teacher ratings were used (29.4 vs. 24.5). This Culture × Rater interaction was significant ($p < .05$). Also, for the rural sample, Chinese children were rated 5 points higher than Americans when parent ratings were used (31.3 vs. 26.0) but 10 points higher when teacher ratings were used (39.0 vs. 29.4). However, the Culture × Rater interaction for the rural sample was not significant, perhaps owing to the smaller n (105 from each country) and the large variance in the American rural TRF sample (SD = 29.70). Like the Thai findings, these findings suggest that contrasts between home and school settings are greater in developing countries than in countries such as the U.S. and Holland, where the CBCL and TRF Total Problem scores are more similar to each other (Achenbach et al., 1987a, 1987b).

A question can be raised about whether Chinese parents are used to questionnaires and to answering them candidly. The Chinese parents were told that all data were coded by number. We also found no significant differences between the Total Problem scores for CBCLs completed independently by parents (M = 27.7 for urban sample; 31.9 for rural) versus those administered by teachers and principals (M = 19.0 for urban sample, n = 5; 31.1 for rural, n = 75).

The cross-cultural validity of the CBCL cross-informant syndromes has been supported by exploratory and confirmatory factor analyses of CBCLs for 4,674 clinically referred Dutch children (DeGroot, Koot, & Verhulst, 1994). In the absence of factor analytic data for Chinese children, we calculated Cronbach's (1951) alpha coefficient of internal consistency for the syndrome, Internalizing, and Externalizing scores separately in the Chinese and American samples. Alpha was significantly lower in the Chinese samples among all four groups (urban CBCL and TRF, and rural CBCL and TRF). Out of 10 comparisons in each sample, eight were significant for the rural CBCL sample, five for both the urban TRF and rural CBCL samples, and three for the urban CBCL sample. Given that 2 out of 10 differences in each sample could be significant by chance and that some

alphas were attenuated by very small total scale scores, the following general trends were observed: Alphas were significantly lower in two out of four Chinese samples on the Attention Problems, Anxious/Depressed, and Somatic Complaints syndromes. It is interesting to note that Anxious/Depressed and Somatic Complaints also had lower alphas for Greek children than American children (MacDonald et al., in press).

Gender Differences

Clear trends emerged regarding gender differences. Among urban children scored by both parents and teachers and rural children scored by teachers, boys displayed significantly higher scores on the Attention Problems, Delinquent Behavior, and Aggressive Behavior syndromes, and on Externalizing. However, these effects were all small. Across urban and rural samples in both cultures, boys scored higher on TRF Total Problems, whereas girls scored higher on the CBCL Somatic Complaints syndrome. Also, in parent and teacher reports for the urban sample, Chinese boys scored higher than both Chinese girls and American boys and girls on the Delinquent Behavior syndrome and Total Problems. This was the only interaction effect that exhibited some consistency across either urban and rural samples or parent and teacher reports.

The gender differences are consistent with those found in the studies of Thai children (Weisz et al., 1989, 1987). In the studies of both Thai and Chinese children, most of the significant gender differences reflected higher scores for boys than girls, and these differences concerned externalizing types of behavior (e.g., Aggressive and Delinquent Behavior, and Attention Problems). Gender differences in this study were consistent with those found in previous studies of Chinese children (Ekblad, 1989; Xin et al., 1992), with analogous differences between boys and girls on other measures. They are also consistent with gender differences found in several other countries (Achenbach, Hensley, Phares, & Grayson, 1990).

CONCLUSIONS

Despite the enormous differences between American and Chinese societies, parents and teachers reported quite similar rates of problems for both urban and rural children in the two countries. Furthermore, the degree of agreement between parents and teachers was remarkably similar, as indicated by correlations of .36 and .29, which were also very close to the correlation of .27 found between parent and teacher ratings in meta-analyses of many studies (Achenbach et al., 1987a). In addition, gender

differences in problem scores for Chinese children were similar to those found in CBCL data for other cultures.

None of the eight syndrome scales, Internalizing, Externalizing, or Total Problem scores showed significant nationality differences that were consistent across parent and teacher ratings in both urban and rural samples. Only one specific problem item, *Prefers being with younger children*, showed a significant nationality difference that was consistent across all four comparisons, with the tendency for Chinese children to obtain higher scores accounting for 6 to 8% of the variance. This difference clearly warrants considerable confidence. Other differences were consistent within parent or teacher ratings or within urban or rural samples. Yet, in view of the major differences between Chinese and American societies, the number and size of differences were remarkably small. Furthermore, despite some significant nationality differences in other comparisons with the U.S. CBCL and TRF data, the overall range of variation has been quite small and the ranking of item scores has been quite similar across diverse nationalities. It thus appears that empirically based assessment methodology for obtaining data on children's problems from parents and teachers is similarly applicable in very different societies.

REFERENCES

Achenbach, T. M. (1991a). *Integrative guide for the 1991 CBCL/4-18, YSR, and TRF profiles*. Burlington: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1991b). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington: University of Vermont, Department of Psychiatry.

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

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

Achenbach, T. M., Hensley, V. R., Phares, V., & Grayson, D. (1990). Problems and competencies reported by parents of Australian and American children. *Journal of Child Psychology and Psychiatry, 31*, 265-286.

Achenbach, T. M., Howell, C. T., Quay, H. C., & Conners, C. K. (1991). National survey of competencies and problems among 4- to 16-year-olds: Parent reports for normative and clinical samples. *Monographs of the Society for Research in Child Development, 56*(3, Serial No. 225).

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987a). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*, 213-232.

Achenbach, T. M., Verhulst, F. C., Baron, G. D., & Akkerhuis, G. W. (1987b). Epidemiological comparisons of American and Dutch children: I. Behavioral/emotional problems and competencies reported by parents for ages 4 to 16. *Journal of the American Academy of Child and Adolescent Psychiatry, 26*, 317-325.

Achenbach, T. M., Verhulst, F. C., Edelbrock, C., Baron, G. D., & Akkerhuis, G. W. (1987c). Epidemiological comparisons of American and Dutch children: II. Behavioral/emotional

problems reported by teachers for ages 6 to 11. *Journal of the American Academy of Child and Adolescent Psychiatry, 26,* 326-332.

Brown, J. S., & Achenbach, T. A. (1995). *Bibliography of published studies using the Child Behavior Checklist and related materials: 1995 edition.* Burlington: University of Vermont, Department of Psychiatry.

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

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297-334.

DeGroot, A., Koot, H. M., & Verhulst, F. C. (1994). The cross-national generalizability of the CBCL Cross-Informant Syndromes. *Psychological Assessment, 6,* 225-230.

Ekblad, S. (1989). Stability in aggression and aggression control in a sample of primary school children in China. *Acta Psychiatrica Scandinavica, 80,* 160-164.

Hollingshead, A. B. (1975). *Four factor index of social status.* Unpublished manuscript. Yale University Department of Sociology, New Haven.

Lambert, M. C., Knight, F., Taylor, R., & Achenbach, T. M. (1994). Epidemiology of behavioral and emotional problems among children of Jamaica and the United States. *Journal of Abnormal Child Psychology, 22,* 113-128.

MacDonald, V. M., Tsiantis, J., Achenbach, T. A., Motti-Stefanidi, F., & Richardson, S. C. (1995). Competencies and problems reported by parents of Greek and American children, ages 6-11. *European Child and Adolescent Psychiatry, 4,* 1-13.

Matsuura, M., Okubo, Y., Kojima, T., Takahashi, R., Wang, Y.-F., Shen, Y.-C., & Lee, C. K. (1993). A cross-national prevalence study of children with emotional and behavioral problems—A WHO collaborative study in the Western Pacific Region. *Journal of Child Psychology and Psychiatry, 34,* 307-315.

McConaughy, S. H., Stanger, C., & Achenbach, T. M. (1992). Three-year course of behavioral/emotional problems in a national sample of 4- to 16-year-olds: I. Agreement among informants. *Journal of the American Academy of Child and Adolescent Psychiatry, 31,* 932-940.

McGee, R., Silva, P. A., & Williams, S. (1984). Behavior problems in a population of seven-year-old children: Prevalence, stability, and types of disorder—a research report. *Journal of Child Psychology and Psychiatry, 25,* 251-259.

Minde, K. K. (1975). Psychological problems in Ugandan school children: A controlled evaluation. *Journal of Child Psychology and Psychiatry, 16,* 49-59.

Olweus, D. (1975). *Development of a multi-faceted aggression inventory for boys.* Bergen: Institute of Psychology, University of Bergen.

Olweus, D. (1984). Stability in aggressive and withdrawn, inhibited behavior patterns. In R. M. Kaplan, V. J. Konecni, & R. W. Novaco (Eds.), *Aggression in children and youth* (pp. 104-137). The Hague: Nijhoff.

Rutter, M., Tizard, J., & Whitmore, K. (1970). *Education, health and behavior.* London: Longman.

Sakoda, J. M., Cohen, B. H., & Beall, G. (1954). Test of significance for a series of statistical tests. *Psychological Bulletin, 51,* 172-175.

Stanger, C., Fombonne, E., & Achenbach, T. M. (1994). Epidemiological comparisons of American and French children: Parent reports of problems and competencies for ages 6-11. *European Child and Adolescent Psychiatry, 3,* 16-28.

Tseng, W.-S., Kuotai, T., Hsu, J., Jinghua, C., Lian, Y., & Kameoka, V. (1988). Family planning and child mental health in China: The Nanjing survey. *American Journal of Psychiatry, 145,* 1396-1403.

Weisz, J. R., 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,* 98-109.

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

Weisz, J. R., Suwanlert, S., Chaiyasit, W., Weiss, B., Achenbach, T. M., & Walter, B. R. (1987). Epidemiology of behavioral and emotional problems among Thai and American children: Parent reports for ages 6 to 11. *Journal of the American Academy of Child and Adolescent Psychiatry, 26,* 890-897.

Xin, R., Chen, S. K., Tang, H. Q., Lin, X. F., & McConville, B. J. (1992). Behavioral problems among preschool age children in Shanghai: Analysis of 3,000 cases. *Canadian Journal of Psychiatry, 37,* 250-258.

Zhang, H. (1988). Psychological measurement in China. *International Journal of Psychology, 23,* 101-117.