A Multimethod Study of Problem Behavior among Thai and American Children in School: Teacher Reports versus Direct Observations

John R. Weisz
University of California, Los Angeles

Wanchai Chaiyasit
Child Mental Health Center, Bangkok, Thailand

Bahr Weiss
Vanderbilt University

Karen L. Eastman
University of California, Los Angeles

Elizabeth W. Jackson
University of North Carolina, Chapel Hill

WEISZ, JOHN R.; CHAIYASIT, WANCHAI; WEISS, BAHR; EASTMAN, KAREN L.; and JACKSON, ELIZABETH W. *A Multimethod Study of Problem Behavior among Thai and American Children in School: Teacher Reports versus Direct Observations.* CHILD DEVELOPMENT, 1995, **66**, 402–415. Previous literature describes Thai children as unusually polite, deferent, and behaviorally restrained. Yet, in a recent study employing teacher reports, Thai children were reported to show many more behavior problems than American children. Such a finding may reflect culture-linked differences in the perspective of Thai versus American teachers. To explore this possibility, we used trained observers to conduct direct observations of Thai and American children's school behavior, and we obtained teacher reports on the same children. Observational results were precisely the opposite of previous and present teacher-report findings: Observers reported twice as much problem behavior and off-task behavior in American children as in their Thai age-mates. This pattern may reflect Thai-U.S. differences in teachers' style, societal values and practices, even child temperament. The findings support the value of direct behavior observation in cross-national research on child problems.

In efforts to transcend a "monocultural science" (Kennedy, Scheirer, & Rogers, 1984), researchers have begun building a base of data on child development and behavior in diverse cultures. The field of developmental psychopathology has been a part of this zeitgeist, with researchers generating data on children's problem behavior in cultures around the world, including North America (Achenbach & Edelbrock, 1983), South America (Montenegro, 1983), the Caribbean (Lambert, Weisz, & Knight, 1989), Australia (Achenbach, Hensley, Phares, & Grayson, 1990), Asia (Weisz, Suwanlert, Chaiyasit, Weiss, Achenbach, & Walter, 1987), Europe (Achenbach, Verhulst, Baron, & Akkerhuis, 1987), and Africa (Weisz, Sigman, Weiss, & Mosk, 1993).

The data for most of this research have consisted of reports on child problems from individuals who are personally involved in

We dedicate this study to the memory of our friend, Frank Bostancic. The investigation was supported by a grant from the National Institute of Mental Health (5 R01 MH38240). We thank Kimberly Cane, Janet Clarke, Mongkol Lugkam, Meredith Mullins, Valerie Proffitt, Somsong Suwanlert, and Ampai Thong ngen for their help in various aspects of the study. Correspondence regarding the study should be addressed to John Weisz, Department of Psychology, Franz Hall, UCLA, 405 Hilgard Avenue, Los Angeles, CA 90024-1563.

[*Child Development*, 1995, **66**, 402–415. 0009-3920/95/6602-0012$01.00]

the lives of the youngsters being studied. In nearly all such studies, these individuals are the children's parents (e.g., Achenbach et al., 1990) or teachers (e.g., Weisz, Suwanlert, Chaiyasit, Weiss, Achenbach, et al., 1988). Certainly, parents and teachers have important perspectives on the behavior of the children they see so regularly, and information from these sources is valuable. On the other hand, it is probably unwise to rely exclusively on such reports, particularly in research that compares samples from different cultures. As has been noted elsewhere (e.g., Weisz, 1989; Weisz, Suwanlert, Chaiyasit, Weiss, et al., 1987), adults' reports on children's behavior and problems may be influenced by cultural milieu in a number of ways; societal values and expectations about child development may color such adult judgments as whether a particular child has shown a particular behavior or problem more than the average child, which behavior is unusual or abnormal, and whether certain behavior has occurred more often than is appropriate for children of a particular age. Such influences may lead to an incomplete or distorted picture of cross-national differences in child behavior.

A possible case in point can be found in research on child behavior in Thailand and the United States. To explain the problem, we need a brief primer on Thai culture and child socialization. The population of Thailand is 95% Buddhist; children are exposed to Buddhist training at home, school, and temple. Consistent with Thai Buddhist precepts, children are reared to be nonaggressive, obedient, and respectful of others—particularly authority figures (e.g., teachers) and others who are older than they—and to avoid expressing anger or other strong emotions (Gardiner, 1968; Gardiner & Suttipan, 1977; Moore, 1974; Sangsingkeo, 1969; Suvannathat, 1979; Suwanlert, 1974; Weisz, 1989). Children are taught to strive for *krengchai*, an attitude of humility and self-effacement intended to avoid disturbing or inconveniencing others (Phillips, 1965; Suvannathat, 1979). The attitude is embodied in the *wai*—a respectful bow with hands pressed together in a prayerful position—with which social interactions in Thailand begin and end. As conveyors of both information and societal values, teachers are accorded special honor in Thai society; parents defer to teachers' judgments about their children, children are taught to obey and honor their teachers, both parents and children use honorific terms (e.g., *achaan*) to refer to teachers, and the entire profession is celebrated on National Teachers Day. Taken together, the literature suggests that Thai children—compared to, say, American children—should show unusually high levels of self-control and deference, and unusually low levels of problem behavior, particularly in the presence of their teachers at school. However, research thus far has found just the opposite.

Weisz, Suwanlert, Chaiyasit, Weiss, Achenbach, et al. (1988), relying on teacher reports of child behavior, found that Thai primary school children were rated as showing much more problem behavior—including more conduct problems—than their American counterparts. It is certainly possible that this finding reflects real Thai-U.S. behavioral differences. However, the findings run counter to the literature noted above and clash with our own impression that Thai children are more orderly, attentive, and well-behaved in their classes than are American children. Such considerations raise the possibility (noted in Weisz, Suwanlert, Chaiyasit, Weiss, Achenbach, et al., 1988) that Thai teachers are stricter than Americans in their expectations and standards for children—that is, that Thai teachers' views on what is average, appropriate, and/or "normal" child behavior lead them to set lower thresholds for problem identification than the thresholds set by American teachers. If this were the case, Thai teachers might well report higher levels of problem behavior than would American teachers even if both groups of teachers were observing the same child behavior at school, and possibly even if Thai children were actually better-behaved or less problem-prone than American children.

One way to explore the plausibility of such alternative explanations is to collect structured observations of child behavior in Thailand and the United States, using observers who (*a*) have been trained to follow well-specified rules in their reports and (*b*) have not developed personal relationships with the children that might color their reports of the behavior they observe. This was the main objective of the present study. In developing the observational method for this study, we faced a number of challenges raised by our emphasis on problem behavior and by the cross-national nature of the project. One challenge was that some of the most

clinically significant behaviors of children are low in base rate and thus unlikely to show much variability during the observational periods that are typically available to observational researchers. A second challenge, applicable to cross-national research, was the impracticality of training a team of observers and transporting them from country to country to do the research; we tried videotapes but found that they could not capture individual child behavior or classroom activity as fully as live observers could. A third challenge was that observers were required who were fluent in the two languages involved and who had some familiarity with customs, conventions, slang expressions, gestures, and the like, in the two cultures. A fourth challenge was that the list of behaviors to be observed needed to include problems of clinical significance in the two countries.

The study was designed to address each of these challenges. Because many clinically significant problems are low in base rate (challenge 1), we trained the observers to identify more than 100 specific problem behaviors and then focused our primary analyses on problem summary scores. Because we needed observers who were fluent in the language and familiar with the culture of the children they observed, but we also needed at least one observer who was involved in the observations in both countries (challenges 2 and 3), we used an observer who was well versed in the language and customs of both countries, and this observer served as the reliability standard for both the American and Thai observer teams. Finally, to insure that our sample of child behaviors to be observed included problems that are clinically significant in the two cultures (challenge 4), we drew from previous research with child clinical samples in both countries (Weisz, Suwanlert, Chaiyasit, & Walter, 1987) and from expert clinician opinions to assemble the final list of behaviors.

The study included attention to the possible impact of gender. In Thailand, boys appear to receive more rigorous training than girls in Buddhist principles and ideals, including the ideals of nonaggression, quiescence, and self-control. For example, boys but not girls are expected to live in a temple for a period of priesthood during their youth. Is such gender-differentiated training associated with a different pattern of gender differences in Thailand than in the United States? Previous Thai-U.S. research (e.g., Weisz, Suwanlert, Chaiyasit, Weiss, et al., 1987; Weisz, Suwanlert, et al., 1993) has not found such differential differences, but the observational methodology used here provided a fresh opportunity to test the possibility.

So, the study took the following form. Teams of trained observers in Thailand and the United States, with one shared team member, coded multiple problems of elementary school children in classroom settings. Individual children were the focus of the observations, and the child (not the class) was the unit of analysis. We then examined the observed prevalence of the problems as a function of culture (Thai vs. United States) and gender. To fill out the picture of Thai and American child behavior in classroom settings, we also coded and analyzed the children's on-task versus off-task behavior. Finally, to provide a comparison of observer reports and teacher reports for the sample of children, we obtained standardized problem reports from teachers on most of the children for whom we had observational data. This made it possible to assess problem reports (on problem items that appeared on both the observer list and the teacher list) as a function of source (i.e., observer vs. teacher), culture, and the theoretically important source × culture interaction.

Method

Subjects and Design of the Study

The sample included 144 children, 36 in each cell of a 2 × 2 (culture × gender) design; thus, culture and gender were orthogonal factors. The children were all pupils in elementary schools. To minimize the chance that results might reflect the idiosyncrasies of any one school, we drew children from multiple schools in each country—six in Thailand and five in the United States. All the Thai children were ethnic Thai; the American sample, reflecting the diversity of the population, was 68% Caucasian and 31% African-American. Within the U.S. sample, we checked for Caucasian versus African-American differences in off-task ratings and total problem scores from the observational measure (see below) and for interactions involving race in analyses of Over- and Under-controlled problems and observer versus teacher differences (see below). The only significant race effect was found for off-task ratings; these are discussed in the relevant section of the Results (below).

Children in the full sample ranged in age from 5 to 11, with an overall mean of

8.13 (SD = 1.93). We tested for group differences, via a 2 × 2 (culture × gender) GLM ANOVA on age; the analysis revealed no significant main effects or interactions (all $ps > .50$). Moreover, a nonsignificant chi-square value for the (7 × 2) yearly age level × culture table, $\chi^2(6, N = 144) = 1.14$, $p > .5$, revealed no Thai-U.S. difference in age distribution.

We sought an SES classification system that could be used to check for group differences; none of the five SES systems proposed by various investigators for Thailand had validity data or coding instructions detailed enough to permit us to apply the system to our data. So, we applied Hollingshead's (1975) nine-level SES scheme (9 = highest SES) for parent occupations to our Thai and U.S. samples. The Thai mean should be interpreted with caution; it reflects use of an American system in a culture for which it was not designed. A 2 × 2 (culture × gender) ANOVA of SES scores showed only a main effect of culture, $F(1, 137) = 12.19$, $p < .001$; the U.S. mean (5.70) was higher than the Thai mean (4.31), reflecting a well-known difference between "developed" and "developing" countries.

Classes, Teachers, and Conditions during Data Collection

In the schools observed in both countries, children stay with their primary (homeroom) teacher during most of each school day, and they change homerooms and primary teachers each calendar year. There were about 35 students in most of the Thai classes; some U.S. classes contained 30 children, but most were smaller. Observations and teacher reports were collected for the study after children had been with their primary teachers for 2 months or more. Thai teachers were all ethnic Thai, and 87% were female; 76% had bachelor's degrees, 19% teacher training or vocational certificates (i.e., less than a bachelor's degree), and 5% had master's degrees. American teachers were 80% Caucasian, 20% African-American, and 95% were female; 90% had bachelor's degrees, and 10% had master's degrees. In both Thailand and the United States, classes involved a mixture of lecturing, questions to and from the students, and projects for groups of students to work on together. Most classrooms in both countries were laid out with combinations of group tables, work stations, and individual, paired, or joined desks; school days in both countries included some time at desks and some time at tables or work stations, or sitting on the floor with groups. It is possible that Thai teachers showed a more authoritative and directing style in interactions with their pupils than did American teachers, but we did not systematically assess teacher style.

Observational Form and Observational Procedure

The observers used an observation form developed for this study and patterned in part after the Direct Observation Form (DOF) of the Child Behavior Checklist (see Achenbach, 1986). The procedure involved time sampling, with one observational form used for each 10-min episode. The form included:

1. *On/Off-task ratings.*—The form included 10 boxes, with "on" and "off" written inside each. The observer's in-the-ear tone sounded after each minute, at which point the observer circled "on" if the child was on-task and "off" if not; when the last box was filled, the 10-min observational episode had ended.

2. *Individual problem ratings.*—The form listed 111 individual problem behavior items (e.g., makes odd noises, argues, teases, physically isolates self from others); beside each item there was space for a rating of 0 (not observed), 1 (slight or ambiguous occurrence), 2 (definite occurrence with moderate intensity and less than 3 min duration), or 3 (definite occurrence with strong intensity or greater than 3 min duration).[1] All of the 96 specific behaviors listed in the Achenbach (1986) Direct Observation Form were included in the present observation form, although some Achenbach items were modified in an effort to clarify the observers' task. For example, to distinguish between Achenbach items "9 Doesn't sit still, restless, or hyperactive" and "13 Fidgets," we changed item 9 to "Doesn't sit still [body trunk movements]" and item 13 to "Fidgets [arm, hand, leg, or foot movements, including fiddling

[1] Using the same four-point rating scale, we also had observers make *prosocial behavior* ratings, using four categories: *Helping or assisting, sharing or giving, praising or encouraging,* and *other efforts to benefit others.* Classroom activities appeared to offer little opportunity for prosocial behavior. Ratings were 0 for 95.0% (for "helping or assisting") to 99.5% (for "other efforts") of all observational sessions, and chance-corrected interobserver reliabilities were, accordingly, not acceptable.

with hair or objects in hands]." In addition to the Achenbach-based items, we added items based on problems listed by Thai parents in a survey of clinic referral problems (see Weisz, Suwanlert, Chaiyasit, & Walter, 1987) and problems nominated as important by Thai clinicians.[2] For example, the Thai emphasis on social harmony appeared to be reflected in such items as "69 Unwilling to share things with others," and "109 Unfair to peers," "16 Impolite, rude, or verbally abusive to peers," and "32 Gets bullied by other children." The Thai emphasis on social appropriateness was seemingly reflected in "108 Clings to peers, or too dependent on peers," "111 Acts too old for his/her age," and "70 Risky or unsafe behavior." The full list of 111 problem items was included in the reliability checks and main analyses reported below. Observers may have been aided in their task by the apparently low base rate of many problems; 47 of the items received a nonzero rating in fewer than 1% of all observations in both countries (17 of these had no nonzero ratings).

Each child was observed for a minimum of four 10-min episodes; the range was 4–15, with a mean of 5.69 sessions per child. Observational scores (e.g., for Total problems) were computed as an average per 10-min episode, so that all available observational data could be used. Children were selected randomly from among those whose parents had given permission, with the constraint that we maintain an age × gender balance. All observations occurred during regular academic class activities; such nonacademic activities as lunch and recess were excluded. For each child, observation times were randomly selected from those class times school staff made available, with the constraint that no child be observed more than once on the same day (this constraint was not applied to reliability assessment observations—see below). This randomization had the effect of scattering observations throughout the day, morning and afternoon, in both Thailand (8:00 A.M.–3:30 P.M.) and the United States (8:30 A.M.–3:00 P.M.). The observation periods were signaled by in-the-ear tones. Immediately after each 10-min observation episode, the observer filled in one copy of the observation form for the child.

Reliability Data on the Observational Procedure

The U.S. observer team consisted of five American observers plus one Thai observer who served as a reliability standard (RS) in both countries; the RS was a clinical psychologist who was fluent in English and Thai, had received his academic training in clinical psychology in the United States, and had worked in clinical and school settings with both American children in the United States and Thai children in Thailand. The RS also served as a member of the observer team for Thailand, which included two additional Thai observers. We calculated interobserver reliabilities separately for the American observers versus the RS, and for the Thai observers versus the RS, using Intraclass Correlation Coefficients (ICCs) between pairs of observers—that is, RS versus other; to aid in interpretation of the ICCs, we also report Pearson *r*s. For the U.S. observers, the number of observational episodes per pair ranged from 9 to 34, with a mean of 16 and a median of 12; for the Thai observers, all observations were done in trios, and there were 47 observational episodes. Overall, 12% of the total observational time was observed for reliability purposes.

For On/Off-task ratings, ICCs for the U.S. observers ranged from 0.39 (low outlier, one observer pair with only nine shared observations) to 0.97, with a median of 0.82 and weighted mean of 0.78. Corresponding Pearson *r*s ranged from 0.44 to 0.97, with a median of 0.82 and weighted mean of 0.79. On/Off reliability for the Thai observers was reflected in an ICC of 0.94 for RS versus observer A and 0.95 for RS versus observer B (ICC for the two non-RS observers: 0.90), and Pearson *r*s of 0.95.

Total Problem score (i.e., sum of all ratings across all problem items) ICCs for the U.S. observers ranged from 0.46 (same outlier pair as above, based on nine observations) to 0.76, with a median of 0.66 and weighted mean of 0.66; corresponding Pear-

[2] In addition to the individual problem ratings, there was space at the end of the form for overall ratings (0, 1, 2, or 3) for "overcontrolled or internalizing problem behavior," "undercontrolled or externalizing problem behavior," and "overall level of problem behavior"; these were included to provide a check on whether the individual problem ratings provided more reliable or valid information than that which could be obtained through simple global ratings. These global ratings did indeed prove unreliable, indicating that observation of the component individual problem items was needed for empirically acceptable observational research of this sort.

son *r*s ranged from 0.43 to 0.78, with a median of 0.77 and weighted mean of 0.71. The Total Problem score ICCs for Thai observers were 0.64 for RS versus A and 0.76 for RS versus B (ICC for A vs. B was 0.62), with a mean Pearson r of 0.75. Thus, the reliabilities all appeared to be acceptable, for both the U.S. observer team and the Thai observer team. The ICCs for On/Off-task were somewhat higher, and the ICCs for Total Problems were somewhat lower, than the corresponding ICCs reported by Reed and Edelbrock (1983) for Achenbach's (1986) DOF, used with an exclusively American sample.

Because we assessed reliability by having multiple observers present for some sessions, it seemed useful to test whether the number of observers present during a session might have influenced children's problem scores. We checked, separately for the Thai and U.S. samples, by computing the correlation between number of observers present and total problem score. We did the calculations for the U.S. sample once using actual number of observers (1, 2, or 3, $r = -0.16$, $df = 97$) and once collapsing to 1 versus >1 observers ($r = 0.14$, $df = 97$). In the Thai sample, only one analysis was needed because the number of observers was either one or three; r for the Thai sample -0.01 ($df = 78$). In all instances, the correlation between child problem score and number of observers was nonsignificant.

Teacher Report Data

Teachers of the children in our Thai and American samples were asked to provide reports of the children's problem behavior. American teachers filled in the problem portion of the Child Behavior Checklist—Teacher Report Form (TRF; Achenbach & Edelbrock, 1986). This is a list of 118 problem items (e.g., argues a lot, cries a lot, disturbs other pupils), each of which is to be rated 0 ("not true of the pupil"), 1 ("somewhat or sometimes true of the pupil"), or 2 ("very true or often true of the pupil"). Thai teachers filled in the problem portion of the Thai Youth Checklist (TYC; Weisz, Suwanlert, Chaiyasit, Weiss, Achenbach, et al., 1988). This consists of the same 118 problem items listed on the TRF, followed by 24 additional problem items based on clinic surveys in Thailand, and the input of Thai clinicians. (Details of instrument development and the three-wave translation–back translation process are provided in Weisz, Suwanlert, Chiayasit, Weiss, Achenbach, et al., 1988.) For analyses involving teacher reports in the present study, we focused on the subset of 86 problem items which were present on both teacher checklists and on the list of problems coded by observers.

Results

We analyzed the different observational measures via 2 × 2 (culture × gender) ANOVAs. In addition to reporting F and p values, we interpreted all significant effects using Cohen's (1988) criteria for magnitude: ANOVA effects are considered *small* if they account for 1%–5.9% of the variance, *medium* if 5.9%–13.8%, and *large* if >13.8%.

On-Task/Off-Task Scores from the Observational Measure

We conducted a 2 × 2 (culture × gender) ANOVA of On/Off-Task scores; these were computed as the arcsine-transformed proportion of off-task ratings across the 1-min intervals in each observational period. The ANOVA yielded only one significant effect, a main effect of culture, $F(1, 140) = 53.94$, $p < .0001$. This effect accounted for 27.38% of the variance in On/Off-Task scores; thus, it was a large effect, by Cohen's (1988) criteria. Means are shown in Table 1. Thai children were off-task in 10% of the observations (SD = 0.11); American children were off-task in 23% of the observations (SD = 0.13), that is, more than twice as often as Thai children. Note that neither the gender nor the gender × culture effect was significant. For comparison purposes, note that studies using the Achenbach (1986) DOF have reported percent off-task means ranging from 11% (Reed & Edelbrock, 1983) to 22% (McConaughy, Achenbach, & Gent, 1988—22% for classroom observations only; off-task scores were lower for lunch and recess observations) for nonclinical samples of American children.

Within the U.S. sample, we found that African-American children were more often off-task than Caucasian children (means: 30% vs. 21%), $F(1, 69) = 7.44$, $p < .01$. Thus, we reran the ANOVA reported in the previous paragraph, including only Caucasian children; the results resembled those described above, with a main effect of culture, $F(1, 117) = 34.86$, $p < .0001$, accounting for 22.59% of the variance. Thai children were off-task in 10% of the observations, American children in 21%.

Total Problem Scores from the Observational Measure

Total problem scores (i.e., sum of all problem item ratings for each 10-min obser-

TABLE 1

MEANS (and SDs) FOR OBSERVER-REPORT AND TEACHER-REPORT MEASURES

	% OffTask	ObsTot	ObsOvr	ObsUndr	TRFTot	TRFOvr	TRFUndr
United States	.23	11.62	.01	.20	15.65	.30	.37
	(.13)	(5.21)	(.02)	(.10)	(14.78)	(.36)	(.43)
Boys	.26	12.92	.01	.22	19.08	.26	.51
	(.14)	(5.52)	(.01)	(.11)	(17.29)	(.36)	(.51)
Girls	.21	10.33	.01	.17	12.77	.33	.26
	(.12)	(4.61)	(.02)	(.09)	(11.82)	(.37)	(.31)
Thai	.10	5.17	.01	.08	31.06	.69	.58
	(.11)	(2.21)	(.02)	(.05)	(21.82)	(.49)	(.50)
Boys	.11	5.54	.01	.08	34.17	.70	.68
	(.11)	(2.20)	(.01)	(.05)	(22.74)	(.46)	(.57)
Girls	.09	4.81	.01	.07	27.94	.68	.48
	(.10)	(2.19)	(.02)	(.04)	(20.70)	(.53)	(.39)

NOTE.—Total problem scores based on observer reports (ObsTot) and teacher reports (TRFTot) represent the sums of all ratings on individual problem items; because observers used a 0–3 scale and teachers a 0–2 scale, the total scores are not directly comparable without transformation. To permit comparison between observer-report means for Overcontrolled and Undercontrolled problems (ObsOvr and ObsUndr) and between teacher-report means for Overcontrolled and Undercontrolled problems (TRFOvr and TRFUndr), these means are all based not on sums of ratings but on the average rating per item.

vation) were subjected to a square root transformation because of skewness in the distribution. The 2 × 2 (culture × gender) ANOVA of these scores yielded a significant main effect of culture, $F(1, 140) = 111.80$, $p < .0001$. The effect accounted for 43.24% of the variance in problem scores, thus easily qualifying as a *large* effect in Cohen's (1988) system. The raw mean problem score was 5.17 (SD = 2.21) for Thai youngsters and 11.62 (SD = 5.21) for their American counterparts. Thus, as was true of the On/Off-Task scores, American children showed problem scores that were more than twice as severe as those of Thai children.

For comparison purposes, we redid the ANOVA, this time including only those 96 observational items derived from Achenbach's (1986) Direct Observation Form. We again found a significant main effect of culture, $F(1, 136) = 113.10$, $p < .0001$, an effect accounting for 42.65% of the variance. Total problem score means were 4.44 for Thai children and 10.05 for American children. Thus, the findings were quite similar to those based on the full set of 111 problem items. Means of 4.5 (Reed & Edelbrock, 1983) and 3.5 (McConaughy et al., 1988) have been reported previously for nonclinical American samples with Achenbach's (1986) DOF.

Undercontrolled and Overcontrolled Problems on the Observational Measure

In an effort to shed some light on which type of problem was most involved in the observational findings reported above, we focused on the distinction between Overcontrolled and Undercontrolled problems (see introduction, above). Because the observations had taken place in school classrooms, we grouped problems into Over- and Undercontrolled types according to whether they loaded on Internalizing versus Externalizing factors, respectively, in principal components analyses of the Teacher Report Form of the Child Behavior Checklist (reported in Achenbach & Edelbrock, 1986).[3] There were 16 TRF Overcontrolled items that also appeared in our observational form (e.g., sad, worries, withdrawn); 27 TRF Undercontrolled items which also appeared in our observational form (e.g., argues, fidgets, swears).

We carried out a 2 × 2 × 2 (culture × gender × problem type) repeated-measures ANOVA of the two observational scores, with problem type (Over- vs. Undercon-

[3] We used the 1986 TRF factors rather than the Achenbach (1991) integrative factors for parent, teacher, and youth self-reports, to (*a*) maintain a focus on factors specific to school-based teacher ratings and (*b*) facilitate comparison with our previous Thai-U.S. teacher report comparison (Weisz, Suwanlert, Chaiyasit, Weiss, Achenbach, et al., 1988), which relied on the 1986 TRF factors.

trolled) as the within-groups factor; because the number of Over- and Undercontrolled items differed, we used as dependent variable the mean score per item, averaging across 16 and 27 items for Over- and Undercontrolled, respectively; the range was thus 0–3 for both scores. The analysis yielded a very highly significant main effect of problem type, $F(1, 140) = 948.98$, $p < .0001$; Undercontrolled problems were observed much more often than Overcontrolled (means: 0.14 and 0.01 on the 0–3 scale). This effect accounted for 76.61% of the variance and was thus large by Cohen's (1988) standards. The only other significant effect was a culture × problem type interaction, $F(1, 140) = 22.95$, $p < .0001$; the effect was small, accounting for 1.85% of the variance. Breaking the interaction down, we found that the problem type effect was highly significant for both the Thai sample, $F(1, 71) = 319.39$, $p < .0001$, and the U.S. sample, $F(1, 71) = 658.54$, $p < .0001$, and, of course, the two effects ran in the same direction. Viewing the interaction from the other direction, we found that the effect of culture was nonsignificant for Overcontrolled problems (U.S. and Thai means were both 0.01 on the 0–3 scale) but highly significant for Undercontrolled problems (U.S. mean 0.20; Thai mean 0.08), $F(1, 142) = 94.28$, $p < .0001$ (a large effect, accounting for 39.90% of the variance). Thus, the large culture effect for total problem scores reflects, in part, the fact that American children showed substantially higher rates of Undercontrolled problems than Thai children, but the two national groups did not differ in their rates of Overcontrolled problems. We did not recompute this set of analyses for Achenbach items only, because Over- and Undercontrolled scores were already based on Achenbach items.

Source Effects: Observational Data versus Teacher Reports

Finally, we undertook an analysis intended to clarify the relation between the present observational findings and the findings obtained with teacher reports in an earlier study by Weisz, Suwanlert, Chaiyasit, Weiss, Achenbach, et al. (1988). In that study, Thai youngsters were reported to show much higher levels of problem behavior than were their American age-mates, when Thai and American teachers filled in Thai and English teacher reports on the youngsters (mean total problem scores: 30.9 vs. 19.3; cf. U.S. means of 22.3 for 6–11-year-old boys and 15.1 for 6–11-year-old girls in Achenbach & Edelbrock, 1986, app. C). The observational findings reported in the present study run in precisely the opposite direction. It is possible that the difference between the findings of the two studies resulted from differences in the groups of children studied, in the schools selected for study, or in other factors unrelated to real differences in the behavior of Thai versus American youngsters. To rule out these extraneous explanations, one needs to have both observational data and teacher reports from the same samples of children. As noted above, we arranged to obtain such data, obtaining teacher reports via parallel Thai and American instruments; we were able to obtain such teacher reports for all 72 of the Thai children and for 57 (i.e., 79%) of the 72 American children.

This made it possible to conduct a 2 × 2 × 2 (culture × gender × source [i.e., observer report vs. teacher report]) ANOVA on Total Problem Scores, summing across those 86 problem items that were common to both the teacher form and the observational form; the 86 items included Overcontrolled and Undercontrolled problems, as noted above, plus other problems (e.g., clings to adults, clumsy, destroys own things, unliked) which could not be clearly classified into either category based on previous principal components analyses. To produce comparable scores for the TRF (0–2 scale) and the observational measure (0–3 scale), we converted the observational measure to a 0–2 scale by multiplying scores by ⅔ (we considered using standard scores, but this would have set means for both measures at 0, thus eliminating any source effects). Our analysis yielded a main effect of source, $F(1, 125) = 167.20$, $p < .0001$ (a large effect, accounting for 22.75% of the variance); scores were higher for teacher reports than for observer reports (means: 21.10 and 4.28). The main effect seems likely to have resulted partly from the fact that teachers had had much more exposure to the children than had the observers.

The most interesting and theoretically significant finding was the culture × source interaction, shown in Figure 1, $F(1, 125) = 48.29$, $p < .0001$ (a medium effect, accounting for 6.57% of the variance). Teacher ratings showed a highly significant Thai-U.S. difference, with Thai children (27.59) showing much higher problem scores than American children (mean 12.91), $F(1, 127) = 30.38$, $p < .0001$ (a large effect, 19.30% of the variance). (For comparison purposes, prorating scores yielded Thai and U.S.

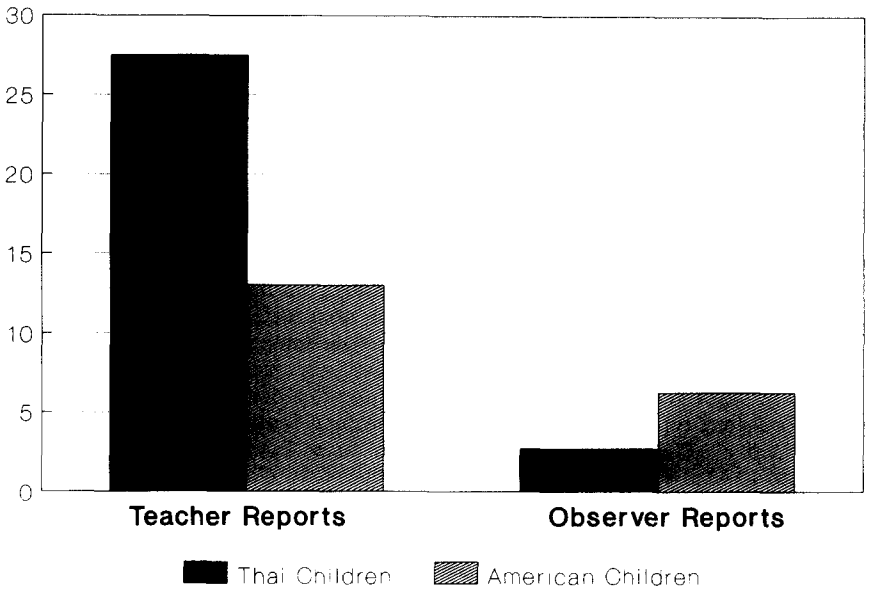

FIG. 1.—Total problem scores of Thai and American children, as reported by the children's teachers and by trained observers.

means of 22.1 and 13.8 in Weisz, Suwanlert, Chaiyasit, Weiss, Achenbach, et al.'s [1988] teacher report study, and U.S. TRF means of 16.0 for boys and 10.8 for girls in Achenbach & Edelbrock's [1986] TRF manual [app. C].) Observer ratings also showed a highly significant Thai-U.S. difference, but in precisely the opposite direction, with Thai children (mean 2.73) showing much *lower* problem scores than American children (mean 6.25), $F(1, 127) = 92.94$, $p < .0001$ (also a large effect, 42.26% of the variance). To summarize this striking effect, when teachers did the rating, Thai children scored twice as high as American children in problem behavior; but when trained observers rated the same children, in the same classes, on the same problem items, the American youngsters scored twice as high as their Thai counterparts.

We recomputed the Observer versus Teacher analyses using only the observational items that were based on Achenbach's Direct Observation Form; observation data were transformed to a 0-1-2 scale, to match the TRF scale. Again, the 2 × 2 × 2 (culture × gender × source) ANOVA yielded a significant main effect of source, $F(1, 125) = 147.51$, $p < .0001$; it accounted for 32.04% of the variance. Mean problem scores were 24.25 for teachers, 4.29 for observers. The culture × source interaction was also significant once again, $F(1, 125) = 32.70$, $p < .0001$, accounting for 7.10% of the variance. The pattern of means resembled that shown in Figure 1. Once again, observer ratings yielded problem scores twice as high for U.S. children as for Thai youngsters (means 6.25 vs. 2.73), whereas teacher ratings yielded problem scores twice as high for the Thai sample as for the American sample (means 31.05 vs. 15.64). The culture effect was significant for both observer and teacher ratings (both $ps < .0001$). And the source effect was significant for both the Thai and American samples (both $ps < .0001$).

Teacher × Observer Correlations

As might be expected from the preceding findings, teachers and observers did not show very high levels of agreement in their ratings of the children. To assess teacher-observer correlations we again focused on only those items present in the rating forms for both informants. For Undercontrolled problems, the teacher-observer correlation was 0.33 in the United States ($df = 56$), 0.19 in Thailand ($df = 71$; z for difference = 0.99, p = N.S.). For Overcontrolled problems, the teacher-observer correlation was 0.13 in the United States, 0.17 in Thailand ($z = 0.82$, p = N.S.). For total problems, the teacher-observer correlation was 0.32 in the United States, 0.15 in Thailand ($z = 0.99$, p = N.S.). Correlations between TRF and Achenbach DOF total problem scores (but not Over- and Undercontrolled), reported elsewhere for mixed clinical-plus-nonclinical samples of

boys (Achenbach & Edelbrock, 1986; Reed & Edelbrock, 1983), have ranged from 0.44 to 0.51 (possibly enhanced by the increased behavioral variability of the mixed clinical and nonclinical samples).

Teacher-Report Findings for Over- and Undercontrolled Problems

Given the emphasis in previous Thai-U.S. research on Overcontrolled and Undercontrolled problems, we tested for nationality and gender differences in teacher reports for the two problem types. We first carried out a 2 × 2 × 2 (problem type × culture × gender) repeated-measures ANOVA, with problem type as the repeated-measure factor. The analyses yielded two significant interactions: Problem type × culture, $F(1, 125) = 3.90$, $p = .05$, and problem type × gender, $F(1, 125) = 7.39$, $p < .01$. The problem type × culture interaction reflected the fact that, although teachers rated Thai children significantly higher than American children on both Over- and Undercontrolled problems, the Thai-U.S. difference was greater for Overcontrolled problems (mean rating per item: 0.69 vs. 0.30; $p < .0001$) than for Undercontrolled problems (means 0.58 vs. 0.37; $p < .01$). Scores for Over- and Undercontrolled problems did not differ significantly within either national sample. The problem type × gender interaction largely reflected the fact that boys were rated higher than girls on Undercontrolled problems (means 0.61 vs. 0.38; $p < .01$) whereas boys and girls had virtually identical Overcontrolled scores (means 0.51 vs. 0.52; $p =$ N.S.). The Over- versus Undercontrolled difference was significant among girls ($p < .05$) but not among boys.

Checking the Impact of SES on Findings

Given the mean SES difference between the Thai and American samples (see above), we tested whether our findings might have been influenced by SES. We recomputed all the ANOVAs reported above, but with SES covaried. Controlling for SES did not change any of the findings in a substantial way; all effects that were previously significant remained significant and within the same p-level category (e.g., at $p < .0001$), and all previously nonsignificant effects remained nonsignificant, with SES controlled.

Most Common Observer-reported and Teacher-reported Problems in Thai and American Samples

Results reported thus far have come from purely quantitative analyses. To complement these findings with information of a somewhat more qualitative sort, we present in Table 2 the 12 most frequently noted problems in the Thai and U.S. samples, based on our observational data and teacher reports. The table reveals numerous similarities in the types of problems reported by observers for Thai and American children, with nine of the top 12 observer-reported problems among Thai children also in the top 12 for American children. The overall pattern suggests that observers in the United States were seeing problems similar to those seen in Thailand, but that U.S. observers were seeing more frequent displays of those problems. By contrast, teacher reports in the two countries seemed to reflect a greater disparity in the types of problems reported; this is reflected in the fact that only four of the top 12 teacher-reported problems among Thai children were in the top 12 for the American sample.

The impression given by Table 2 was reinforced by correlational analyses involving the 86 problem items that were common to both the teacher and observer forms. The mean correlation between Thai observational scores and U.S. observational scores, across the 86 items, was 0.87 ($df = 110$, $p < .0001$), whereas the mean correlation for teacher ratings was 0.57 ($df = 118$, $p < .0001$); the difference between the two coefficients was significant, $z = 5.13$, $p < .00001$.

Discussion

As anticipated, the observational findings revealed a pattern of Thai-U.S. differences quite different from the pattern shown in teacher reports, thus lending support to the possibility that teacher reports are colored by culture-linked values and expectations about child behavior. Teachers in previous research (Weisz, Suwanlert, Chaiyasit, Weiss, Achenbach, et al., 1988) and in the present study, reported levels of problem behavior that were twice as high for Thai children as for their American age-mates. By contrast, direct observations in the present study showed the reverse pattern: Twice as much problem behavior among American children as among Thai children, a difference so large that it accounted for 43% of the variance in problem scores. Moreover, direct observations also showed that American youngsters were off-task 23% of the time, compared to 10% for Thai children.

Although the present observational findings were derived from U.S. and Thai children, it is tempting to compare them to

TABLE 2

TWELVE MOST COMMON PROBLEMS AMONG THAI AND AMERICAN CHILDREN AS REPORTED BY TRAINED OBSERVERS AND BY TEACHERS

REPORTS BY TRAINED OBSERVERS

American	%	Thai	%
1. Easily distracted (6)	73.8	1. Fidgets (10)	57.6
2. Fidgets (10)	68.2	2. Easily distracted (6)	45.2
3. Doesn't sit still (8)	65.5	3. Doesn't sit still (8)	25.8
4. Doesn't concentrate/pay attention (5)	46.7	4. Picks nose, skin (42)	24.67
5. Talks when shouldn't (74)	36.7	5. Eats/drinks nonfood (24)	22.5
6. Sucks thumb, hand (25)	25.1	6. Yawns (75)	19.0
7. Yawns (75)	23.5	7. Sucks thumb, hand (25)	16.8
8. Picks nose, skin (42)	22.9	8. Talks when shouldn't (74)	11.4
9. Talks too much (94)	16.6	9. Apathetic, unmotivated (44)	6.3
10. Shows off/clowns (50)	13.3	10. Doesn't concentrate/pay attention (5)	6.3
11. Sings/talks to self (73)	12.2	11. Argues (2)	6.3
12. Teases (14)	11.6	12. Shows off/clowns (50)	5.2

REPORTS BY TEACHERS

American	%	Thai	%
1. Talks too much (93)	50.9	1. Can't concentrate/pay attention (8)	70.8
2. Talks out of turn (53)	45.6	2. Messy work (72)	69.4
3. Hurt when criticized (81)	42.1	3. Fails to carry out assigned tasks (100)	66.7
4. Disturbs peers (24)	40.4	4. Dependent on adults (11)	66.7
5. Easily distracted (78)	38.6	5. Sulks a lot (88)	65.3
6. Messy work (72)	38.6	6. Easily distracted (78)	63.9
7. Demands attention (19)	36.8	7. Fails to finish things she or he starts (4)	63.9
8. Argues a lot (3)	35.1	8. Acts without thinking (41)	61.1
9. Fails to finish things she or he starts (4)	32.1	9. Hurt when criticized (81)	59.7
10. Self-conscious, easily embarrassed (71)	31.6	10. Apathetic, unmotivated (60)	59.7
11. Misdeeds with no guilt (26)	31.6	11. Nervous, high-strung (45)	59.7
12. Disrupts class (67)	29.8	12. Can't sit still (10)	58.3

NOTE.—Numbers in parentheses are the item numbers from the observation form (top half of table) and Teacher Report Form (bottom half). Numbers in percent column for Trained Observers indicate the percent of observations for which nonzero ratings were recorded for the particular item; numbers in percent column for Reports by Teachers indicate the percent of children for whom nonzero ratings were recorded for the particular item.

findings reported by Stevenson and Stigler (1992) in their comparisons of youngsters in the United States and two other Asian nations, Japan and Taiwan. Compared to children in the two Asian countries, Stevenson and Stigler (1992) found that "American children spent the least amount of time actually engaged in academic activities in academic classes" (p. 147). Moreover, "American children spent more time out of their seats, talked more to their peers at inappropriate times, and engaged in other inappropriate activities to a greater degree than did Chinese and Japanese children. For example, American fifth-graders were out of their seats nearly 20 percent of the time; Chinese and Japanese fifth-graders, less than 5 percent" (p. 148). The findings summarized by Stevenson and Stigler (1992) suggest that at least some of our results may resemble patterns seen in other Asian versus American comparisons.

Our observational data do not provide direct evidence on reasons for the Thai-American differences, but several possible reasons warrant attention. First, Thai instructional procedures and classroom man-

agement practices appear to be somewhat more strict, authoritative, and controlling than the American norm. Thai practices, like the Japanese and Taiwanese procedures described by Stevenson and Stigler (1992), may foster higher levels of classroom discipline than American schools, thus stimulating more on-task activity and better behavioral control among Thai than American children. A second possibility is that broad cultural differences between Thailand and the United States may be reflected in the findings discussed here. For example, as noted above (and see Gardiner, 1968; Gardiner & Suttipan, 1977; Moore, 1974; Suvannathat, 1979; Weisz, 1989), Thai children are reared to be deferent, compliant, and respectful of others, particularly those who are older and those in positions of authority; and teachers are treated with special respect throughout the society, addressed with honorific terminology, and supported by parents when difficulties arise with their children (cf. similar phenomena described by Stevenson & Stigler, 1992). A third possibility is that Thai children differ from American children in temperament (see Weisz, Suwanlert, et al., 1993), manifesting a more behaviorally and emotionally inhibited style that makes them more attentive and less likely to show behavior problems. Data from other Asian groups at least raise this possibility (see e.g., Freedman, 1979). Thus, the Thai-U.S. differences observed in the study may reflect such diverse influences as instructional styles and classroom management practices, cultural traditions (including respect for teachers) and child temperament. These causal possibilities warrant attention in future research.

Such causal speculations must be tempered by the fact that Overcontrolled problems played essentially no role in our observational findings, presumably because such problems are not very evident to external observers during brief observation periods. Previous Thai-U.S. comparative findings using parent reports (Weisz, Suwanlert, Chaiyasit, & Walter, 1987; Weisz, Suwanlert, Chaiyasit, Weiss, et al., 1987; Weisz, Suwanlert, et al., 1993) indicate that Thai youngsters may be especially susceptible to Overcontrolled problems. It is possible, then, that a measure more sensitive to children's internal processes might have shown higher levels of at least this particular form of problem behavior in Thai than American school children.

The most valuable general lesson of this study may be that much can be learned by directly observing children's behavior, rather than relying solely on the reports of informed but untrained informants, such as teachers. Our observational findings ran directly counter to the findings of Weisz, Suwanlert, Chaiyasit, Weiss, Achenbach, et al. (1988) based on Thai and American teachers' reports of their elementary school pupils' behavior. In the 1988 study, Thai teachers reported significantly higher levels of Overcontrolled, Undercontrolled, and Total Problems among Thai children than American teachers reported for American children. In principle, it is possible that the difference between the earlier teacher report findings of Weisz, Suwanlert, Chaiyasit, Weiss, Achenbach, et al. (1988) and the present findings resulted from between-study differences in the schools or samples involved, or the fact that the teacher reports involved a somewhat different collection of problem items than did the observer reports. However, one analysis reported above casts doubt on these artifactual possibilities. We compared observers' reports and teachers' reports for the same children (i.e., children from the present study), from the same schools, and using the same collection of items (i.e., only the 86 items that were included in both the teacher form and the observer form). The results showed quite clearly that teacher reports for the present sample generated higher problem ratings for Thai than American children, whereas the reports of trained observers generated precisely the opposite picture for the same children. The results suggest that Thai teachers may evaluate their students' behavior using more exacting standards than American teachers do; perhaps Thai teachers are so accustomed to high levels of respect, deference, obedience, and behavioral control by their pupils that they rate as serious behavior problems what would be seen as normal or minor by American teachers. Alternatively, perhaps Thai teachers are more sensitive than American teachers to signs of behavioral or emotional problems in children. Finally, Thai teachers may have been more responsive than American teachers to demand characteristics of the assessment procedure—for example, they may have been especially conscientious about reading each checklist item carefully and trying to remember all of each child's relevant past behaviors.

Whose reports should we rely on—observers or teachers? As some of the preceding discussion implies, we place some-

what greater confidence in the observers' reports than in the teachers' ratings. Observers were trained to use common standards for reporting on child behavior, even across the two different cultures; and we have data supporting the observers' interrater reliability. On the other hand, it is certainly possible that teachers' day-to-day interactions with their pupils give them a more comprehensive picture of, and deeper insight into, the children's behavior than that which could be achieved by outside observers during a series of 10-min sessions. Our observers focused on classroom behavior, whereas teachers are also privy to children's lunch time and recess behavior, and at least one study of American children (McConaughy et al., 1988) points to lower levels of off-task and problem behavior at lunch and recess than in class; our exclusive focus on classroom behavior ruled out detection of possible culture × situation interactions within the school setting.

To make the observer versus teacher debate even more complex, we must stress that there is no true "right answer" to the question of whether a particular child "has" a particular problem. As we have discussed previously (Weisz, 1989; Weisz, Suwanlert, Chaiyasit, Weiss, Walter, et al., 1988), the concept of a child problem, like the concept of child psychopathology generally, involves two components that are very difficult to disentangle: Actual child behavior and the perception of that behavior through the lens of some observer—a teacher, a trained observer, a parent, or the child him- or herself. "Actual behavior" and "perceived behavior" may be inextricably linked in reports by any informant; in a sense, it is precisely the blend of actual child behavior and culturally conditioned perceptions that forms the essence of "child problems" in the first place.

Such reasoning, combined with the present findings, underscores the need to view the child's behavior through more than one window. This need is especially pronounced in cross-national research on child behavior and child problems. The perspectives of parent, teachers, trained observers, and the children themselves may all be important sources of information. As suggested above, one might argue that trained observers can provide the most unbiased and reliable information on child behavior, because such observers can be trained and held to standards of reliability that may not apply to parents, teachers, or children. Yet, observer reports are almost certain to be based on a less complete sample of child behavior than reports of the other three types of informants. This is especially important when the target behavior of interest involves psychological problems, many of which are quite low in base rate. For example, potentially significant problem behavior such as vandalism, suicidal talk or attempts, fire-setting, and stealing may never occur in the presence of observers; information about such problems may be accessible only to informants who share significant portions of the child's life space—informants such as parents, teachers, or the children themselves.

Thus, for many reasons, some empirical and some conceptual, no single source of information on child behavior is likely to provide a "magic bullet" or final "right answer." Instead, our closest approach to the "truth" about child behavior and child problems is apt to involve a triangulation of reports from multiple sources, with each offering a different base of information and a different perspective on the child.[4] What the present study suggests is that the reports of trained observers will be an important part of that mix of information and perspectives.

References

Achenbach, T. M. (1986). *The Child Behavior Checklist—Direct Observation Form.* Burlington, VT: University of Vermont Department of Psychiatry.

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

Achenbach, T. M., & Edelbrock, C. S. (1983). *Manual for the Child Behavior Checklist and Revised Child Behavior Profile.* Burlington, VT: University of Vermont Department of Psychiatry.

Achenbach, T. M., & Edelbrock, C. S. (1986). *Manual for the Teacher's Report Form and teacher version of the child behavior profile.* Burlington, VT: University of Vermont Department of Psychiatry.

Achenbach, T. M., 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., Verhulst, F. C., Baron, G. D.,

[4] For a description of recent technology for integration of data on child problem behavior from multiple sources, see Achenbach (1991).

& Akkerhuis, G. W. (1987). 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.

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

Freedman, D. G. (1979). Ethnic differences in babies. *Human Nature*, **2**, 36–43.

Gardiner, H. W. (1968). Expression of anger among Thais: Some preliminary findings. *Psychologia*, **11**, 221–228.

Gardiner, H. W., & Suttipan, C. S. (1977). Parental tolerance of aggression: Perceptions of preadolescents in Thailand. *Psychologia*, **20**, 28–32.

Hollingshead, A. B. (1975). *Four-factor index of social status*. Unpublished manuscript, Yale University.

Kennedy, S., Scheirer, J., & Rogers, A. (1984). The price of success: Our monocultural science. *American Psychologist*, **39**, 996–997.

Lambert, M. C., Weisz, J. R., & Knight, F. (1989). Over- and undercontrolled clinic referral problems of Jamaican and American children and adolescents: The culture-general and the culture-specific. *Journal of Consulting and Clinical Psychology*, **57**, 467–472.

McConaughy, S. H., Achenbach, T. M., & Gent, C. L. (1988). Multiaxial empirically based assessment: Parent, teacher, observational, cognitive, and personality correlates of child behavior profile types for 6- to 11-year-old boys. *Journal of Abnormal Child Psychology*, **16**, 485–509.

Montenegro, H. (1983). *Salud mental del escolar: Estandarizacion del inventario de problemas conductuales y destrezas sociales de T. Achenbach en ninos de 6 a 11 anos*. Santiago, Chile: Centro de Estudios de Desarrollo y Estimulacion Psisocial, Roman Diaz 26, Of. 63.

Moore, F. J. (1974). *Thailand: Its people, its society, its culture*. New Haven, CT: HRAF Press.

Phillips, H. P. (1965). *Thai peasant personality: The patterning of interpersonal behavior in the village of Bang Chan*. Berkeley: University of California Press.

Reed, M. L., & Edelbrock, C. (1983). Reliability and validity of the Direct Observation Form of the Child Behavior Checklist. *Journal of Abnormal Child Psychology*, **11**, 521–530.

Sangsingkeo, P. (1969). Buddhism and some effects on the rearing of children in Thailand. In W. Caudill & T. Y. Lin (Eds.), *Mental health research in Asia and the Pacific* (pp. 286–295). Honolulu: East-West Center Press.

Stevenson, H. W., & Stigler, J. W. (1992). *The learning gap: Why our schools are failing and what we can learn from Japanese and Chinese education*. New York: Summit Books.

Suvannathat, C. (1979). The inculcation of values in Thai children. *International Social Science Journal*, **31**, 477–485.

Suwanlert, S. (1974). Some personality characteristics of Thai students. In W. P. Lebra (Ed.), *Youth, socialization, and mental health in Asia and the Pacific* (pp. 75–84). Honolulu: University Press of Hawaii.

Weisz, J. R. (1989). Culture and the development of child psychopathology: Lessons from Thailand. In D. Cicchetti (Ed.), *The emergence of a discipline: Rochester symposium on developmental psychopathology* (Vol. **1**, pp. 89–117). Hillsdale, NJ: Erlbaum.

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., & Walter, B. R. (1987). Over- and Undercontrolled referral problems among Thai and American children and adolescents: The *wat* and *wai* of cultural differences. *Journal of Consulting and Clinical Psychology*, **55**, 719–726.

Weisz, J. R., Suwanlert, S., Chaiyasit, W., Weiss, B., Achenbach, T. M., & Eastman, K. L. (1993). Behavioral and emotional problems among Thai and American adolescents: Parent reports for ages 12–16. *Journal of Abnormal Psychology*, **102**, 395–403.

Weisz, J. R., Suwanlert, S., Chaiyasit, W., Weiss, B., Achenbach, T. M., & Trevathan, D. (1988). 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. A. (1987). Epidemiology of behavioral and emotional problems among Thai and American children: Parent reports for ages 6–11. *Journal of the American Academy of Child and Adolescent Psychiatry*, **26**, 890–898.

Weisz, J. R., Suwanlert, S., Chaiyasit, W., Weiss, B., Walter, B., & Anderson, W. W. (1988). Thai and American perspectives on Over- and Undercontrolled child behavior problems: Exploring the threshold model among parents, teachers, and psychologists. *Journal of Consulting and Clinical Psychology*, **56**, 601–609.