

Studying the "Referability" of Child Clinical Problems

John R. Weisz
University of North Carolina at Chapel Hill

Bahr Weiss
Vanderbilt University

Child clinical problems may differ in their power to evoke clinic referral, and this "referability" of problems may also differ with the gender or culture of the child who manifests them. The authors propose (a) a method for the study of such differences and (b) a new statistic, the *referability index* (RI). RI reflects the frequency with which a child problem stimulates clinic referral, adjusted for its prevalence in the general population. RI was assessed across 118 child problems as a function of problem type (overcontrolled vs. undercontrolled), child gender, and culture (U.S. vs. Thailand). Problems of both types were more referable in girls than boys; undercontrolled problems were more referable than overcontrolled in the U.S. but, surprisingly, not in Thailand. The findings shed new light on gender and culture differences and illustrate the empirical and heuristic potential of the RI statistic.

The study of child psychopathology is actually the study of two somewhat distinct phenomena: the behavior of children and the responses of adults to that behavior. Descriptive epidemiologic research can address the first phenomenon, telling us the frequency with which various kinds of child behavior and child problems occur. However, such research needs to be complemented by research on the second phenomenon, to help us understand the significance and consequences of various forms of child behavior within their social context. Because children rarely consider themselves "psychologically disturbed" and rarely refer themselves for treatment, it is generally adults in a particular child's society who determine whether that child's behavior constitutes "psychopathology" and whether it warrants intervention. Child problems that adults do not consider serious are not likely to receive clinical attention, even if they are very distressing to the child. Child problems that adults *do* consider serious may well lead to treatment, even if they are not distressing to the child.

Because the adult response so often determines whether a particular child problem will lead to mental health intervention, investigators have begun to study the attitudes and responses of adults to various forms of child "problem behavior" (e.g., Walker, Bettes, & Ceci, 1984; Weisz et al., 1988). A primary goal of such research is often to gauge the extent to which various child problems are likely to be considered serious and to be referred to a specialist for help (i.e., problem *referability*). The referability of different problems may differ from one another, of course, as a function of various characteristics of the

problem (e.g., whether it is an internalizing or "overcontrolled" problem on the one hand or an externalizing or "undercontrolled" problem on the other; see early work on this distinction by Achenbach [1966]). In addition, referability of the *same* problem may differ as a function of child characteristics (e.g., the gender of the child who manifests the problem) or of the culture in which the problem occurs.

Studying referability of child problems is a potentially important enterprise, both theoretically and practically. Consider the following examples: (a) Determining which child problems are high in referability, and which are low, may enrich our understanding of those factors that make adults take child problems seriously. For instance, results of one recent study (Weisz et al., 1988) suggest that parental judgments about how serious and how much in need of treatment a problem is, are related to how troubling or bothersome the behavior is to others. (b) Discovering that a particularly serious problem shows low referability can sensitize us to the need to educate the public, thus enhancing access by children to services they need. (c) Finding differences in the referability of particular problems for girls versus boys may help us identify and address sex stereotypes and biases that can limit access to treatment for boys or girls. (d) Discovering cross-national differences in the referability of particular problems can enhance our understanding of cultural differences and their impact on mental health service delivery. In these and other ways, knowing the referability of various child problems may broaden our knowledge base and enrich our theories of child psychopathology, while enhancing children's access to mental health care.

For these reasons, it is important to the study of child psychopathology that we have accurate information on the referability of child problems. Unfortunately, efforts to generate such information have thus far relied heavily on adults' self-reports (e.g., Likert ratings of how serious particular child problems are, or statements of whether the adults would be likely to refer a child with particular problems for treatment). Such self-reports are useful, but they should not be our only source of data, because they may be subject to possible response biases (e.g., a desire to be cautious, to appear enlightened, or to avoid the

This study was supported by Grant 5 R01 MH38240 from the National Institute of Mental Health, which we acknowledge with appreciation.

We also thank our Thai and American colleagues, Thomas Achenbach, Wannu Wibulswasdi Anderson, Wanchai Chaiyasit, Elizabeth Jackson, Somsong Suwanlert, and Bernadette Walter, for their numerous contributions to this project.

Correspondence concerning this article should be addressed to John R. Weisz, who is now at Department of Psychology, Franz Hall, University of California at Los Angeles, 405 Hilgard Avenue, Los Angeles, California 90024-1563.

appearance of sex bias). Apart from such systematic biases, there is always the possibility that what adults honestly believe they would do (when they answer hypothetical questions on a questionnaire) will differ substantially from what they would actually do when confronted with a real child who has real problems. The challenge for researchers, then, is to find an approach to studying referability that does not rely on adults' self-reports of what they believe or what they would do in hypothetical situations. As a complement to the self-report approach, we propose a method that derives information about referability from the actual behavior of adults confronted with real child problems.

We will illustrate the use of such an approach. In essence, we operationally define referability as the frequency with which a child problem is reported as a presenting complaint at mental health clinics, *relative to* the frequency of that problem in the general population. The choice of this definition is based on the assumption that the frequency with which a problem is reported as a presenting complaint in clinics is primarily a function of (a) the frequency with which it occurs in the general population and (b) the perception that the problem requires and is appropriate for psychological treatment. By adjusting the presenting problem prevalence for the general population prevalence, one is left with the "referability" of a particular problem.

This approach provides a particularly comprehensive *and behavioral* look at referability of child problems. It does not rely on adults' verbal reports of their attitudes, but rather on statistical analysis of the *behavioral consequences* of these attitudes (i.e., of the extent to which referral rates for various child problems over- or underrepresent the prevalence of those problems in the general population).

Beyond the study of specific problems, this approach permits the investigation of whether certain theoretically important child *syndromes* are more referable than others. In our study, we focused on what appear to be the two most frequently identified empirically derived syndromes: *overcontrolled problems* (e.g., anxiety, somatic problems, social withdrawal) and *undercontrolled problems* (e.g., disobedience, fighting, stealing). These two "broadband" syndromes have emerged from more than a dozen independent factor analytic studies of children's behavior problems (see Achenbach & Edelbrock, 1978). There is some evidence that undercontrolled problems predict more serious long-term outcomes for children (see, e.g., Patterson, DeBaryshe, & Ramsey, 1989; Robins, 1979). Recent research using adult self-reports of their attitudes (Weisz et al., 1988) revealed that undercontrolled problems are rated as more serious and worrisome than overcontrolled problems; yet, in that same self-report research, Americans rated the two syndromes as equal in need for referral. Here we probed for differences in referability using the more behavioral assessment approach described above.

In addition to permitting comparisons of the referability of various child problems and syndromes, the approach developed here also makes it possible to investigate demographic factors that may influence referral. One such factor may be child gender. Adult attitudes toward boys and girls may make certain problems differentially referable depending on the gender of the child who manifests the problems. Might an adult

belief that "boys will be boys" make fighting less referable (because it is seen as less serious) among boys than girls? Might an adult perception that girls are moody make bouts of depression or emotional lability less referable for girls than boys? To investigate such possibilities, we explored relations between child gender and referability.

We also explored the possibility that referability may differ with the culture in which child problems occur. We compared two cultures that are said to differ markedly in their orientation toward overcontrolled and undercontrolled child problems, but about which the research evidence thus far is conflicting: Thailand and the United States. Much of the literature on mental health in Thailand suggests that Thai adults, partly because of their Theravada Buddhist orientation, are unusually accepting of quietness, shyness, inhibition, and other forms of *overcontrolled* behavior in children, but that they strongly disapprove of disobedience, aggression, and other forms of *undercontrolled* behavior in children (see, e.g., Gardiner & Suttipan, 1977; Sangsingkeo, 1969; Weisz, 1989). However, a recent study of Thai and American adults' attitudes revealed no reliable differences between the two groups in the perceived relative seriousness or need for treatment of overcontrolled versus undercontrolled child problems (Weisz et al., 1988). Further complicating the picture, another recent study revealed that American children were actually *more* likely than Thai children to be referred to mental health clinics for undercontrolled problems, whereas Thai children were more likely than Americans to be referred for overcontrolled problems (Weisz, Suwanlert, Chaiyasit, & Walter, 1987). These findings, of course, may have been strongly influenced by the actual prevalence of the various problems in the two countries; and the adult attitude findings may have been influenced by the potential biases to which self-report approaches are subject. Both problems are addressed with the present approach to referability assessment; it (a) adjusts for the general population prevalence of the child problems under study and (b) avoids the potential pitfalls of self-report attitude assessment.

In the present study, then, we computed the referability of 118 clinically significant child problems using the calculation methods detailed below. We then examined referability as a function of problem type (overcontrolled vs. undercontrolled), child gender, and culture (Thai vs. American), as well as interactions among these factors.

Method

Two types of samples were included: (a) a general population sample of Thai and American 6- to 11-year-olds and (b) a sample of Thai and American 6- to 11-year-olds who had been referred to mental health and child guidance clinics for treatment of behavioral or emotional problems.

American and Thai General Population Samples

The total general population sample included 960 children, 600 from the United States and 360 from Thailand. Within each culture, the sample included children aged 6-11, half boys and half girls. The American children were drawn from urban, suburban, and semirural environments in the mideastern United States, with census tract and block data used to randomly select families for participation (see

Achenbach & Edelbrock, 1981). Of those families asked, 82.3% took part. The Thai general population sample was obtained from urban, suburban, and semirural environments in Bangkok and four other districts of the country; see Weisz, Suwanlert, Chaiyasit, Weiss, et al. (1987). Because census tract and block data were unavailable in Thailand, we used school districts and schools as the sampling frame (all Thai children are required to complete 7 years of elementary school). Children were sampled from 38 elementary schools (29 public, 9 private, to mirror the proportion of such schools nationwide). From each school we randomly selected grade levels, classes, and one child from each class. Parents (or guardians) of the selected children were then asked to participate; 91.5% of these took part.

All data were gathered through individual interviews. In the United States all interviews were conducted in subjects' homes; Thai interviews were conducted at home or at the child's school, depending on parent preference. In each country, the interviewer read aloud a standardized problem report measure (see below), while the parent followed along on another copy; the interviewer recorded the parent's answers.

In the United States, the problem report measure was the Child Behavior Checklist (CBCL; Achenbach & Edelbrock, 1983). The CBCL includes a list of 118 specific problems (e.g., argues a lot, fears going to school). Parents indicate the degree to which their child shows each problem by ratings of 0, 1, or 2. Test-retest reliability of the CBCL problem reports estimated by intraclass correlation coefficients (ICC) between parent reports at 1-week intervals was 0.95; the ICC for inter-interviewer reliability averaged 0.96 (Achenbach & Edelbrock, 1983). Principal components analyses have revealed broadband overcontrolled and undercontrolled syndromes, as noted above.

Thai interviewers administered the Thai Youth Checklist (TYC). This Thai-language measure is similar in format and content to the CBCL, and the 118 problem items on the CBCL are listed as the initial problem items on the TYC. These 118 problems are the focus of the present analyses. Parent ratings follow the same 0-1-2 rating scale used for the CBCL. Test-retest reliability of the TYC was found to be 0.81 for parents retested after a 1-week interval; interinterviewer reliability was 0.91 (see Weisz, Suwanlert, Chaiyasit, Weiss, et al., 1987).

American and Thai Samples of Clinic-Referred Children

The sample of clinic-referred youngsters included 192 from the United States and 188 from Thailand. Within each culture, the sample included children aged 6-11, half boys, half girls, all drawn from government-funded clinics that serve the general public regardless of ability to pay. The U.S. sample was drawn from clinics in Washington, DC, Charlotte, North Carolina, and rural areas of North Carolina and Tennessee. The Thai sample came from clinics in Bangkok and two large rural provinces (Weisz, Suwanlert, Chaiyasit, & Walter, 1987). Although the Washington, Charlotte, and Bangkok clinics provided primarily urban youngsters, children referred to those clinics who actually lived in rural areas were classified according to their place of residence.

In both countries, cases were drawn from clinic records blindly but with the constraints that there be equal numbers of boys and girls within each national sample and that neither culture nor gender be confounded with age (this was accomplished by filling equal quotas of boys and girls at each yearly age level within each national sample). All clinic data were gathered by trained recorders in the two countries. For all the youngsters sampled, the recorders reviewed the full written report of intake interviews with parents (or guardians), listing verbatim each of the child's problems reported to admitting clinicians during the interview. The two Thai recorders, working independently on 56 cases, achieved good agreement on number of problems reported, $r = 0.96$; the figure for the two U.S. recorders, working on 56 U.S. cases, was $r =$

0.99. The mean number of referral problems was 5.0 in the United States and 4.7 in Thailand (difference = *ns*).

Thai problems were translated into English by two bilingual Thai psychologists, checking each other's work through back translation and retranslation and consulting with a third bilingual expert when disagreements arose. Two trained U.S. coders judged whether each individual Thai or U.S. problem matched any of the 118 CBCL problems, and if so which one. For this coding the problems were written in English, listed individually, and not labeled as to their country of origin. Working independently on a sample of 276 problems from 75 youngsters, the two coders agreed on 96.7% of the problems in their judgments as to whether the problem did or did not have a CBCL equivalent. (More than 80% of the problems listed within each culture were judged to have a CBCL equivalent.) Moreover, for those problems deemed to have an equivalent, the coders agreed 95.1% of the time as to which of the 118 CBCL problems was the appropriate match.

Classifying Problems as Overcontrolled or Undercontrolled

All problems noted for Thai and U.S. children in all the samples were coded as overcontrolled, undercontrolled, or other, on the basis of their loadings in the Achenbach and Edelbrock (1983) principal components analysis of the CBCL for boys and girls aged 6-11.¹ Problems that loaded on the overcontrolled factor for boys or girls or both, but *not* on the undercontrolled factor for either, were classified as overcontrolled; problems were classified as undercontrolled in analogous fashion; remaining problems were labeled *other* (this category included "mixed" problems, that is, those loading on both broadband factors).

Results

Before presenting findings, we describe the analytic procedures.

The Referability Index (RI)

The statistic of primary interest here is what we will call the *referability index* (RI), generated through categorical data modeling procedures known as logit analysis (Agresti, 1984; Feinberg, 1980). A logit is the natural log of an odds ratio (i.e., the log of the probability of being in a particular category divided by the probability of being in any other category); here the categories are the presence or the absence of a particular problem, as reported by parents. Hence, the logit for Problem X in the clinic-referred sample is: $\text{logit } X = \log_N (p/q)$, where p is the probability that Problem X is listed as a referral problem and q is the probability that Problem X is not thus listed. The logit for Problem X in the general population sample is calculated in analogous fashion except that p and q represent the probability that X is versus is *not* reported by parents as occurring in their children.

RI is the logit for the clinical sample minus the logit for the general population sample, divided by 2: $\text{logit (clinic)} - \text{logit}$

¹ Ideally, one would want to classify problems into overcontrolled and undercontrolled categories based on broadband factors identified in both cultures involved. At this time, however, no data are available on the factor structure of child problems in Thailand; so we relied on the U.S. principal components findings of Achenbach & Edelbrock (1983).

(general)/2. Thus computed, RI becomes easily interpretable: If $RI > 0$, then the clinic sample rate is larger than the rate in the general population; if $RI < 0$, then the reverse is true. Thus, the larger the RI, the higher the referability of the problem.

Because RI represents the logit parameter estimate for the effect of sample (i.e., clinic vs. general), it is relatively simple to compare referability across groups (e.g., two different cultures). For example, one may simply include culture as a factor in a logit analysis and inspect the Culture \times Sample interaction. The parameter estimate for this interaction is equal to RI for the first culture minus RI for the second culture divided by 2: $RI(\text{culture A}) - RI(\text{culture B})/2$. Such parameter estimates also are readily interpretable: If $RI > 0$ for a particular problem, then that problem is more referable in the first culture than the second; if $RI < 0$, then the reverse is true.

Interpreting RI values. The procedures outlined above generate an estimate of the *magnitude* of group differences. We can also assess the *reliability* of any such differences. Just as two effect sizes of equal magnitude may not have equal reliability, two between-groups comparisons may yield identical RI differences of unequal reliability. When the proportion of subjects reported to have a particular problem approaches 0 or 1, the transfer of a single count between response groups (i.e., a change in one child's status regarding the presence or absence of the behavior) will have a much larger impact on the RI than when the proportion of subjects reported to have the problem is 0.5. This is true because at the extremes a single count represents a much larger proportion of the smaller group than when the proportion of presence versus absence is closer to 50/50. Hence, differences between logits, the basis for RI comparisons, are less reliable when the groups are distributed near the extremes. To gauge the reliability of any group difference in RI, say, the Culture \times Sample interaction, or $RI(\text{culture})$, one may use the associated probability that a difference of this size would be found by chance, given a true population RI difference of 0.

Although the preceding discussion suggests that the use of probability values with RIs is straightforward, there is a complication: Application of conventional significance criteria (e.g., an effect must have $p < .05$ to be considered significant) would make it impossible to find gender or culture effects for rare problems (e.g., daytime enuresis), regardless of the magnitude of gender or culture differences. This is because, as noted above, very low frequency (and very high frequency) problems inevitably produce unreliable estimates, and comparison of unreliable estimates across groups (e.g., cultures) will also produce unreliable (i.e., "nonsignificant") estimates. For this reason, we present significance tests for group differences (i.e., gender, culture, and problem type) in mean RI across problems. Then we compare the magnitude of RI for individual problems as a function of factors that showed significant effects across problems.

Referability Index for Individual Problems: Validity Check

We first calculated an overall RI for each of the 118 individual problems within the United States, collapsing across child gender. This provided a validity check on the procedure (i.e., a check on whether problems of a seemingly serious nature do in

Table 1
Most and Least Referable Problems in the U.S. Sample, With Referability Index (RI) for Each Problem

Most referable problems		Least referable problems	
Problem	RI	Problem	RI
Vandalism	1.13	Bragging	-2.94
Poor school work	0.87	Teases a lot	-2.62
Runs away from home	0.86	Feels must be perfect	-2.58
Truancy	0.83	Self-conscious	-2.55
Sexual problems	0.78	Allergy	-2.39
Steals outside home	0.77	Fears doing bad	-2.25
Attacks people	0.68	Overtired	-2.12
Suicidal talk	0.49	Too concerned neatness	-2.08
Daytime enuresis	0.46	Likes to be alone	-2.07
Deliberate self-harm	0.46	Obsessions	-2.03
Steals at home	0.32	Prefers older kids	-1.98
Encopresis	0.30	Accident prone	-1.96
Gets in many fights	0.19	Picks nose or skin	-1.95
Destroys others' things	0.14	Lacks guilt	-1.94
Disobeys at school	0.13	Whining	-1.94
Sets fires	0.09	Impulsive	-1.91
Cruel to animals	0.02	Feels unloved	-1.76
Withdrawn	-0.03	Shows off	-1.73
Threatens people	-0.05	Prefers younger kids	-1.69
Lying or cheating	-0.09	Sleeps less than most	-1.67

fact have a higher RI than seemingly less serious problems). Table 1 shows the 20 most referable and the 20 least referable problems, of the 118 total problems studied. The contents of the table appear to support the validity of the RI statistic, in that problems with the highest RI (e.g., vandalism, theft, physical assault, self-harm) generally appear to be relatively serious in their implications for self or others, and problems with the lowest RI (e.g., teasing, perfectionism, overconcern with neatness) do not.

Overall Effects of Culture, Gender, and Problem Type on RI

We next assessed the overall impact of gender, culture, and problem type on RI. We used a $2 \times 2 \times 2$ (Gender \times Culture \times Problem Type) repeated measures analysis of variance (AN-OVA), with RI as the dependent variable² and the problem

² To assess the validity of an analysis of variance (ANOVA) with these referability index (RI) data, we tested the basic ANOVA assumptions of normally distributed data and homogeneity of variance. In general, we found that the data were normally distributed; however, when we tested separately for each of the eight cells, we found one cell that was significantly non-normal (i.e., overcontrolled problems among Thai girls [Shapiro & Wilk, 1965], $W = 0.93$, $p < .005$). We next tested whether variance differed across the levels of Gender, Culture, and Problem Type; here too, only one of these comparisons showed a significant difference: overcontrolled (0.78) vs. undercontrolled (0.41), $F(46, 49) = 1.90$, $p < .05$. To assess whether these two seemingly minor departures from ANOVA assumptions might have biased our p values, we ran a simulation study, generating a 100,000 observation data set with the same three factors as in our actual data (i.e., Gender, Culture, Problem Type) and distributed similarly (i.e., variance, skewness, kurtosis) as our actual data. However, in the simulation data set there were no

items as observations. Because items rather than people were the observations, or “subjects,” our between-subjects and within-subject factors were the reverse of what is usually the case. Gender and Culture were within-subject factors; that is, RIs were computed for each item at each level of gender and culture. Problem type was a between-subjects factor; that is, each problem item was classified as either overcontrolled ($n = 50$ items), undercontrolled ($n = 47$ items), or Other ($n = 21$ items). For this initial analysis, we included only overcontrolled and undercontrolled items, because items in the Other category do not have a clear theoretical or empirical meaning. We complemented this approach with an analysis including *all* items but excluding the Problem Type factor. We report results with both approaches.

There were several significant effects in the initial three-factor analysis: main effects for Gender, $F(1, 95) = 12.50, p < .001$, and Culture, $F(1, 95) = 19.90, p < .0001$, and interactions of Culture \times Gender, $F(1, 95) = 11.53, p < .001$, and Culture \times Problem Type, $F(1, 95) = 17.73, p < .0001$. When we carried out the ANOVA with Problem Type dropped from the model and all problem items (i.e., overcontrolled, undercontrolled, and Other) included as observations, the Culture and Gender main effects and the Culture \times Gender interaction remained highly significant, with no other effects significant.

Gender main effect. Mean RI was higher for girls (-0.901) than for boys (-1.030). Because this main effect was qualified by a Gender \times Culture interaction, we will describe the trend when we describe that interaction.

Culture main effect. RI was higher in the U.S. sample (-0.836) than in the Thai sample (-1.095). Table 2 lists the 20 problems that showed the most pronounced culture differences involving higher referability in the U.S. and the 20 that showed the most pronounced Thai > U.S. differences. We will discuss the table below, under *Culture \times Problem Type interaction*.

Gender \times Culture interaction. The Gender \times Culture interaction, shown in Figure 1, was broken down by general linear models (GLM) simple effects tests on each factor, within both levels of the other factor. These tests showed that the gender effect was significant in the United States, $F(1, 96) = 23.47, p < .0001$ (with girls higher in RI than boys), but not in the Thai sample. Viewing the interaction from the other direction, we

relations between the factors (i.e., the null hypothesis was valid). Next we randomly drew 100 samples from the simulation data set, each containing 47 undercontrolled RI and 50 overcontrolled RI (i.e., same number as our actual sample). For each of the 100 samples we tested the Gender \times Culture \times Problem Type model. If our data satisfied ANOVA assumptions, we would expect that for each parameter (e.g., Culture main effect), random variability would produce five effects significant at 0.05 out of the 100, and the nominal alpha would equal the actual alpha. A larger number of effects in the simulation sample for any parameter would indicate that our nominal alpha (i.e., the alpha produced by the ANOVA) for that parameter was biased (i.e., that there was actually a greater than alpha probability of our results having been produced by chance, rather than true population differences). Each of the four tests significant in our actual data (i.e., Gender, Culture, Culture \times Gender, Culture \times Problem Type) produced exactly 5 of 100 effects significant at alpha .05, indicating that the probability values associated with these estimates were not biased.

Table 2
Problems Showing the Largest Thai–U.S. Differences in Referability, With Parameter Estimate for Each Problem

U.S. > THAI			THAI > U.S.		
Type	Problem	Parameter estimate	Type	Problem	Parameter estimate
M	Sulks	1.13	O	Strange Behavior	-0.67
U	Gets in fights	0.97	O	Fears school	-0.64
U	Swearing, obscene	0.71	O	Headaches	-0.58
M	Mood changes	0.71	O	Problems with eyes	-0.53
O	Constipated	0.70	M	Likes to be alone	-0.51
U	Cruel to animals	0.67	M	Speech problems	-0.51
U	Threatens people	0.65	U	Public sex play	-0.46
O	Unhappy, depressed	0.64	O	Self-conscious	-0.45
U	Lying or cheating	0.60	O	Vomiting	-0.40
U	Sexual problems	0.58	O	Overtired	-0.40
U	Argues	0.56	O	Visual hallucinations	-0.39
M	Unliked by children	0.55	—	Bites nails	-0.35
M	Talks too much	0.55	—	Other physical problems	-0.32
U	Sets fires	0.54	O	Stomachaches	-0.30
—	Doesn't eat well	0.53	O	Nervous, twitches	-0.29
O	Talks suicide	0.53	M	Underactive	-0.26
O	Strange ideas	0.52	M	Acts too young	-0.25
U	Unusually loud	0.46	O	Rashes, skin problems	-0.23
M	Feels guilty	0.45	—	Eats nonfood	-0.23
M	Withdrawn	0.43	U	Cruel to others	-0.20

Note. O = overcontrolled problem, U = undercontrolled, M = mixed (i.e., loads on both syndromes), Dash = no broadband syndrome.

found that the effect of culture was minimally significant for boys (with RI higher for U.S. than for Thai boys), $F(1, 96) = 3.99, p < .05$, but the culture effect was highly significant for girls, with American girls much higher in RI than Thai girls, $F(1, 96) = 30.75, p < .0001$.

Because the male–female difference was significant only in the U.S. sample, we compared, for the U.S. sample only, those problems showing the largest male > female difference in RI to those showing the largest female > male difference. The only noticeable trend was that the female > male list appeared to include a relatively high number of sex-related problems; in addition to 96. Thinks about sex too much, 73. Sexual problems, and 60. Plays with sex parts too much, which are obviously sexual in nature, the list included 52. Feels guilty, 93. Talks too much, and 63. Prefers older children, which also load on the Sex Problems factor for 6- to 11-year-old girls (see Achenbach & Edelbrock, 1983), suggesting an association with sex-related problems.

Culture \times Problem Type interaction. GLM simple effects tests showed that the Culture \times Problem Type interaction, graphed in Figure 2, resulted partly from the fact that the effect of Problem Type was nonsignificant in the Thai sample but significant in the U.S. sample (where undercontrolled problems were more referable than overcontrolled), $F(1, 95) = 7.98, p < .006$. Viewing the interaction from the other perspective, the Culture effect

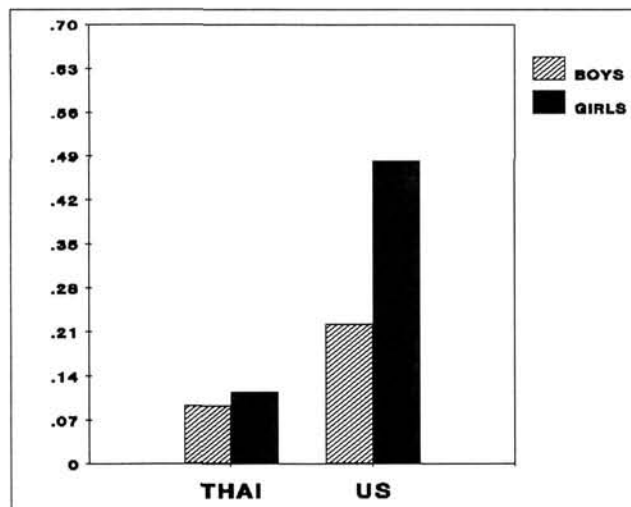


Figure 1. Gender \times Culture interaction: Referability of problems for boys and girls in Thailand and the United States (vertical axis shows mean referability index).

was highly significant for undercontrolled problems (RI higher in the United States than in Thailand), $F(1, 46) = 45.03, p < .0001$, but nonsignificant for overcontrolled.

Table 2 shows the nature of the Culture \times Problem Type interaction: Undercontrolled problems predominate in the list of problems that are more referable in the United States than Thailand (with 9 undercontrolled and only 4 overcontrolled problems), whereas overcontrolled problems predominate in the Thailand > U.S. list (with 11 overcontrolled and only 2 undercontrolled problems). The overall table of Thai/United States versus Over/Under/Other was significant, $\chi^2(2, N = 40) = 7.72, p < .01$. A noteworthy characteristic of the 20 problems in the Thai > United States list is that 10 are explicitly somatic in nature (e.g., headaches, stomachaches, rashes) and 3 have somatic overtones (e.g., eats nonfood).

Effects of Culture and Gender on Ordering of RI

The preceding analyses indicate the extent to which such factors as Gender and Culture were related to the overall level of RI, pooling across problems. These analyses do not, however, indicate whether such grouping factors were related to the relative RI of the items, in comparison with one another. For example, although the main effect for Gender showed that mean RI was higher for girls than boys, it did not reveal whether the ordering of items with respect to RI was different for boys and girls. A significant gender difference in mean RI might or might not involve gender differences in the ordering and distance between individual problem items with respect to RI. We tested for such differences as a function of Gender and Culture (such an analysis is impossible for within-group factors such as Problem Type).

First we assessed the impact of gender on the ordering of the individual problems. For each of the 118 problems, we computed RI twice, once for boys and once for girls. We then computed a Pearson correlation, with the problem items serving as

the matched-pair observations. This produced $r = 0.86 (p < .0001)$, indicating that the item order in RI was highly consistent across gender. We then computed an analogous correlation for the Thai sample versus the U.S. sample; here $r = 0.71 (p < .0001)$ also indicated rather consistent rank orderings across the two cultures. These two findings indicate that Culture and Gender did *not* substantially influence the *relative* referability of problems, compared with one another, but that the two factors instead had their major impact on the overall level of problem referability.

These analyses helped clarify the nature of the Gender and Culture main effects. Next we turned to the Gender \times Culture interaction reported above. Pearson r s were computed between boys and girls separately for the Thai and the U.S. samples, and between cultures separately for boys and girls. Consistency of item ordering between boys and girls was strong within the Thai sample ($r = 0.85, p < .0001$) and within the U.S. sample ($r = 0.84, p < .0001$), indicating that the magnitude of the effect of gender on item order was quite similar across cultures. Item order consistency between the Thai and U.S. samples was marginally stronger for girls ($r = 0.74, p < .0001$) than for boys ($r = 0.62, p < .0001$; z for difference = 1.71, $p < .10$).

Discussion

The findings suggest a number of ways in which the study of referability generally, and the RI statistic in particular, may contribute to our understanding of child psychopathology. First, and most directly, the RI provides a quantitative way of comparing various child problems for their relative "clinical pull" or power to evoke clinic referral. The data in Table 1 suggest that RI values do have considerable clinical validity, with problems such as stealing, vandalism, self-harm, and attacking others showing high RI in the United States. On the other hand, a few potentially serious problems (e.g., feeling unloved, fearing one's own impulses, both of which load on the

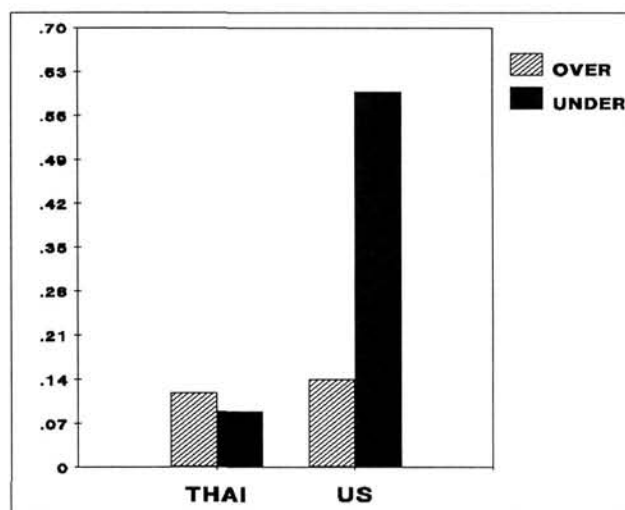


Figure 2. Culture \times Problem Type interaction: Referability of overcontrolled and undercontrolled problems in Thailand and the United States (vertical axis shows mean referability index).

CBCL depression factor for boys and girls, see Achenbach & Edelbrock, 1983) showed low RI. This fact illustrates the potential value of the RI as an indicator of areas where public awareness may need to be sharpened.

The findings also illustrate how the study of referability may shed light on adults' responses to problems in girls versus boys. The results showed problems to be more referable in girls than boys, at least in the U.S. sample. Why then are boys so much more often referred to clinics than girls? Perhaps because the rate (or severity) of actual problem behavior is higher in boys than girls (Graham, 1979). When we compared the problems showing the largest male > female differences in RI in the United States to those problems showing the largest female > male differences in the United States, only one thematic difference was evident: The 20 problems highest on the female > male list included 6 problems that load on the Sex Problems factor (see Achenbach & Edelbrock, 1983). With the exception of sexual preoccupation (see Achenbach & Edelbrock, 1981), sex problems are not generally found to be higher in prevalence among girls than boys; however, the present RI data suggest that when girls do display such problems, they may be more likely to be taken for treatment than are boys who show such problems.

The gender findings also illustrate the fact that assessing referability from data on people's actual behavior may yield a very different picture than surveys in which people give self-reports of their attitudes and beliefs. Weisz et al. (1988) compared Thai and American adults' judgments as to how serious, how worrisome, how unusual, and how likely to improve without treatment various psychological problems were when they occurred in boys versus girls. The findings of that study revealed *no* evidence of a difference in adult attitudes toward psychological problems in boys versus girls. Yet the present RI results include a Gender main effect and a Gender \times Culture interaction, indicating that adults' actual *behavior* does differ as a function of gender, at least in the United States, with American adults more likely to refer girls than boys for problems within the array surveyed here. It is possible that when adults respond to questions about referral *in the abstract* on self-report surveys, they perceive their attitudes as being similar toward boys and girls, but that when their own children manifest *actual* problems, parents are more concerned about, and more likely to refer, their daughters than their sons. It is also possible that issues of social desirability make adults reluctant to give self-reports of different attitudes toward boys than girls. Whatever their proper interpretation, these findings do suggest that the RI approach used here may be a useful means of studying differences in actual adult *behavior* in response to girls' and boys' problems.

The cross-national findings illustrate another kind of difference between self-report attitude findings and behavior-based findings. In the Weisz et al. (1988) study cited above, a culture main effect (with no interactions) indicated much higher levels of concern over *both* overcontrolled and undercontrolled child problems among Americans than among Thais. Yet, in the present study a Problem Type \times Culture interaction indicated that there was no reliable Thai-U.S. difference in the referability of overcontrolled problems, but undercontrolled problems were

significantly more referable in the United States than in Thailand.

Such differences between patterns revealed by self-reported attitudes and patterns evident in actual referral behavior lead us to an important point concerning RI analysis: Referability is overdetermined and certainly reflects more than just the attitudes of referring adults. The referability of a particular problem within a particular culture certainly *may* reflect the extent to which adults are concerned about the problem and believe their child needs help. However, high referability may also reflect numerous other factors, such as, (a) the extent to which a particular problem is thought to require professional mental health intervention, as opposed to, say, remediation within the home; (b) the extent to which professionals are deemed *capable* of resolving the particular problem; (c) the extent to which professionals who specialize in the particular problem are *available* to the child and family; and (d) the extent to which parents are willing to "go public" with a problem that might otherwise be hidden at home (see Lin, Tardiff, Donetz, & Goresky, 1988). Moreover, referability may be influenced by a variety of family demographic and psychological factors ranging from SES and parent education to parental personality and psychopathology (cf. Brody & Forehand, 1986, on the impact of maternal depression on perceptions of child behavior).

Overall, the findings presented here suggest that the RI statistic may provide information of value to mental health professionals and researchers, particularly as a complement to typical forms of epidemiologic data. In addition to knowing the population prevalence of various child problems, there appears to be value in knowing the extent to which the problems stimulate referral and the extent to which the referability of problems depends on such factors as child gender and culture. In fact, one can imagine future epidemiologic studies in which tables include columns of population prevalence statistics paired with parallel RI statistics and in which population group differences in referability as well as prevalence would be a focus of interest.

References

- Achenbach, T. M. (1966). The classification of children's psychiatric symptoms: A factor-analytic study. *Psychological Monographs*, 80(Whole No. 615).
- Achenbach, T. M., & Edelbrock, C. S. (1978). The classification of child psychopathology: A review and analysis of empirical efforts. *Psychological Bulletin*, 85, 1275-1301.
- Achenbach, T. M., & Edelbrock, C. S. (1981). Behavioral problems and competencies reported by parents of normal and disturbed children aged 4-16. *Monographs of the Society for Research in Child Development*, 46(1, Serial No. 188).
- Achenbach, T. M., & Edelbrock, C. S. (1983). *Manual for the Child Behavior Checklist and Revised Child Behavior Profile*. Burlington: University of Vermont.
- Agresti, A. (1984). *Analysis of ordinal categorical data*. New York: Wiley.
- Brody, G. H., & Forehand, R. (1986). Maternal perceptions of child maladjustment as a function of the combined influence of child behavior and maternal depression. *Journal of Consulting and Clinical Psychology*, 54, 237-240.

- Feinberg, S. (1980). *The analysis of cross-classified categorical data* (2nd ed.). Cambridge, MA: MIT Press.
- Gardiner, H. W., & Suttipan, C. S. (1977). Parental tolerance of aggression: Perceptions of preadolescents in Thailand. *Psychologia, 11*, 221-228.
- Graham, P. (1979). Epidemiological studies. In H. C. Quay & J. S. Werry (Eds.), *Psychopathological disorders of childhood* (2nd ed., pp. 185-209). New York: Wiley.
- Lin, T.-Y., Tardiff, K., Donetz, G., & Goresky, W. (1988). Ethnicity and patterns of help-seeking. *Culture, Medicine, and Psychiatry, 2*, 3-13.
- Patterson, G. R., DeBaryshe, B. D., & Ramsey, E. (1989). A developmental perspective on antisocial behavior. *American Psychologist, 44*, 329-335.
- Robins, L. N. (1979). Follow-up studies. In H. S. Quay & J. S. Werry (Eds.), *Psychopathological disorders of childhood* (2nd ed., pp. 483-513). New York: Wiley.
- 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, HI: East-West Center Press.
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*, 591-611.
- Walker, E., Bettes, B., & Ceci, S. (1984). Teachers' assumptions regarding the severity, causes, and outcomes of behavioral problems in preschoolers: Implications for referral. *Journal of Consulting and Clinical Psychology, 52*, 899-902.
- Weisz, J. R. (1989). Culture and the development of child psychopathology: Lessons from Thailand. In D. Cicchetti (Ed.), *Rochester Symposium on Developmental Psychopathology: Vol I* (pp. 89-117). New York: Cambridge University Press.
- Weisz, J. R., Suwanlert, S., Chaiyasit, W., & Walter, B. R. (1987). Over- and undercontrolled referral problems among children and adolescents from Thailand and the United States: 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., & 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*, 890-897.
- Weisz, J. R., Suwanlert, S., Chaiyasit, W., Weiss, B., Walter, B. R., & 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.

Received February 28, 1990

Accepted August 13, 1990 ■