Empirically Based Assessment and Taxonomy of Psychopathology: Cross-Cultural Applications. A Review.

Frank C. Verhulst* and Thomas M. Achenbach**

This paper provides an overview of empirically based assessment and taxonomy, as illustrated by cross-cultural research on psychopathology. The empirically based approach uses standardized assessment procedures to score behavioral and emotional problems from which syndromes are derived by multivariate analyses. Items and syndromes are scored quantitatively to reflect the degree to which individuals manifest them, as reported by particular informants. Although the approach to assessing problems and to constructing taxonomic groupings differs from the ICD/DSM approach, there are no inherent contradictions between either their models for disorders nor the criterial features used to define disorders. Cross-cultural comparisons have yielded relatively small differences in problem rates and syndrome structure, plus considerable similarity in associations of problems with sex and SES, as well as similar correlations between reports by different types of informants. Research on variations in problems in relation to culture, sex, age, SES, and type of informant can contribute to improving both the ICD/DSM and empirically based approaches and to a more effective synthesis between them.

Introduction

We have been asked to present our empirically based approach to assessment and taxonomy in a way that will be helpful to workers in a variety of countries. We will first present the key features of our empirically based approach. This approach has diverse applications to clinical services, training, and research. However, we have chosen to illustrate the approach in terms of cross-cultural applications, which should be of particular interest to the international readership of this journal.

Originally developed to advance our understanding of child and adolescent psychopathology, the empirically based approach has also been extended to young adulthood (Ferdinand et al., in press; Achenbach et al., in press). The main challenge that initially prompted the empirically based approach was the need for a better differentiated nosology for child and adolescent behavioral/emotional disorders (Achenbach, 1966). Although a differentiated nosology for adult disorders had been available since the 19th century (Kraepelin, 1883), very few diagnostic categories were designed for preadult disorders. Furthermore, those few diagnostic categories had not been derived from assessment of representative samples of children or adolescents. To advance communication, clinical services, training, theory, and research, it was necessary to determine what constellations of behavioral/emotional problems actually occurred among troubled youngsters.

As electronic computers became available, multivariate statistical methods were increasingly used to identify constellations of problems that tended to co-occur. Reviews of the early efforts concluded that certain sets of co-occurring problems

*   Sophia Children's Hospital Erasmus, University Rotterdam, Department of Child and Adolescent Psychiatry, Dr. Molewaterplein 60, 3015 GJ Rotterdam, The Netherlands
**   Center for Children, Youth and Families, Department of Psychiatry, University of Vermont, USA.

were repeatedly found, despite wide variation in methodology (Achenbach & Edelbrock, 1978; Quay, 1979). Building on these "first generation" efforts to detect patterns of co-occurring problems, later research tested the replicability of specific syndromes of problems in parents' ratings of multiple clinical samples (Achenbach et al., 1989). Thereafter, cross-informant syndromes were derived that embodied sets of problems found to co-occur in parent, teacher, and self-ratings of children and adolescents referred for mental health or special education services (Achenbach, 1991 a). Because the cross-informant syndromes reflect the common elements of patterns found in ratings of different samples, by different types of informants, on different instruments, they provide a basis for taxonomically representing different kinds of psychopathology.

Features of the Empirically Based Approach

The empirically based approach uses standardized assessment procedures to score the behavioral and emotional problems of large samples of subjects. Multivariate analyses are then applied to the problem scores of clinically referred subjects in order to identify syndromes of co-occurring problems. The scores obtained by an individual can be used to determine how closely that individual's problems resemble each of the empirically derived syndromes. The syndromes thus provide a basis for describing and grouping individuals according to the problems that they manifest. They also provide a basis for targeting interventions, for evaluating changes in response to interventions, and for focusing research and theory on particular kinds of problems.

Comparison with the Nosological Approach of ICD and DSM

There is no inherent contradiction between our empirically based approach to deriving syndromes and the nosological approach embodied in the *International Classification of Diseases* (World Health Organization, 1992) and recent editions of the *Diagnostic and Statistical Manual* (DSM) of the American Psychiatric Association (1980, 1987, 1994). In fact, numerous studies have reported statistically significant associations between our empirically derived syndromes and many clinical variables, in-

cluding DSM diagnoses (e.g., Edelbrock & Costello, 1988; Weinstein et al.,1990; Rey & Morris-Yates, 1992; Gould et al.,1993; Chen et al., 1995).

Similarities to the ICD/DSM Approach

Although our empirically based approach uses quantitative methods to identify co-occurring problems and to score individuals in terms of scales, it does not assume that every disorder merely involves quantitative gradations along a particular dimension. On the contrary, because no one yet knows the true nature and boundaries of many behavioral/emotional disorders of childhood, quantitative methods can help to determine which problems are more effectively conceptualized as categorical, which are more effectively conceptualized as quantitative, and which require a mixture of categorical and quantitative concepts. Much research has already made use of the quantitatively derived syndromes to identify categorical cutpoints that discriminate effectively between subjects who do versus do not meet categorical diagnostic criteria (e.g., Chen et al., 1995). Other research has used cluster analysis to identify types of problem patterns for categorizing individuals according to their overall profiles of syndrome scale scores (Edelbrock & Achenbach, 1980; Achenbach, 1993).

Just as the quantitative derivation of syndromes does not inherently contradict categorical models for disorders, the nosological approach does not inherently contradict quantitative models for disorders. Hypertension, for example, is defined largely in terms of quantitative deviations from norms for blood pressure. Many other quantitative parameters, such as body temperature, heart rate, and white cell count, provide diagnostic criteria for a variety of diseases.

Differences from the ICD/DSM Approach

To advance our understanding of child psychopathology, we need reliable, valid, and generalizable data from both the empirically based and ICD/DSM approaches. Nevertheless, we also need to be aware of some differences between them at this point in their development.

One difference is that the empirically based approach progresses "from the ground up" – that is, it begins with standardized assessment of large samples

of subjects and then uses multivariate analyses to derive patterns of co-occurring problems from the data. These statistically derived patterns can be used as a basis for taxonomic groupings of individuals. This approach has focused mainly on problems for which children are most often referred for mental health services, rather than on low prevalence disorders such as autism.

By contrast, the ICD/DSM approach can be described as progressing "from the top down" – that is, this approach starts with assumptions about the disorders that exist. Committees of experts then formulate criteria for determining whether an individual has a particular disorder. The assumptions and criteria are based on accumulated clinical experience, rather than being statistically derived from standardized assessment of specific samples of subjects. This approach has focused on rare as well as common disorders.

Besides differing in how it derives criteria, our empirically based approach also differs from the ICD/DSM approach in its use of quantification to assess individuals. Rather than requiring present-versus-absent judgments of each criterial feature, the empirically based approach quantifies judgments by using scores such as 0 = *not true* of the subject, 1 = *somewhat or sometimes true*, and 2 = *very true or often true*. Informants base their scores on knowledge of the subject over a specified period, such as 6 months. In addition, each subject's standing on each syndrome scale is quantified by summing the subject's scores on the items comprising the scale. The degree to which subjects are reported to manifest the problems of a particular syndrome is thus measurable in terms of scores ranging from 0 (i.e., no item is true of the subject) to 2 × the number of items on the syndrome (i.e., all items of the syndrome are scored 2, indicating that they are all very true or often true of the subject). Furthermore, the subject's score on the scale can be compared with the scores of peers by converting the raw syndrome score to a percentile or a standard score (such as a T score) based on a particular reference sample.

For example, suppose we wish to determine how 13-year-old Jennifer's score on a particular syndrome compares with the scores of other 13-year-old girls on that syndrome. If a representative sample of girls has been assessed using the same procedure and the same type of informants, the distribution of syndrome scores for that representative sample can be used to determine how high Jennifer's score is in comparison to other girls

of her age. To provide a comparison with "normal" girls, Jennifer's score can be compared with the distribution of scores for girls who are considered normal according to a particular criterion, such as not having been referred for mental health services. Other comparison groups can also be used, such as girls who are judged to need mental health services. Both nonclinical and clinical criterion groups are needed to distinguish between scores that are in the normal versus clinical range. By comparing the scores of nonclinical versus clinical criterion groups, we can identify cutpoints on the distribution of syndrome scores that effectively distinguish between the normal and clinical ranges. To highlight the area of overlap between scores from normal and clinical criterion groups, a borderline clinical range can also be demarcated.

Nonclinical and clinical criterion groups are often used to identify cutpoints on distributions of scores for biomedical measures such as blood pressure. Similar procedures are potentially applicable to the choice of criteria and cutpoints for diagnoses of behavioral/emotional disorders as well. However, the DSM currently defines disorders as being either present or absent, based on present-versus-absent judgments of each criterial feature. If each of the requisite criterial features is judged to be present, an individual is concluded to have the disorder. On the other hand, if an individual is judged to have even one less than the requisite criterial features, then the individual is concluded to be free of the disorder. The ICD is less explicit than the DSM in requiring present-versus-absent judgments about criterial features, but it also implies that most disorders must be judged either present or absent.

From Assessment to Taxonomy (and Back Again)

When mental health workers think about kinds of disorders or groupings of disorders, they are more likely to think in terms of "diagnosis" or "classification" than "taxonomy." However, both diagnosis and classification have meanings that may hamper efforts to develop clearer concepts of children's behavioral/emotional problems, especially for cross-cultural applications. On the one hand, the term "diagnosis" can be confusing because it encompasses (a) *diagnostic processes* – the

gathering of data about cases; (b) *formal diagnoses* – assignment of cases to diagnostic categories, such as those of the ICD and DSM; and (c) *diagnostic formulations* – comprehensive statements about inferred case dynamics. On the other hand, the term "classification" encompasses diverse groupings that may be irrelevant to the intrinsic characteristics of the cases.

Despite being less familiar, the term "taxonomy" emphasizes the goal of identifying patterns that mark important differences between groups of individuals. Within a particular culture, taxonomic distinctions among patterns of problems are essential for improving communication about individual cases; for choosing interventions on the basis of their efficacy in ameliorating particular patterns of problems; for evaluating outcomes; and for advancing research and theory regarding etiology, prevention, treatment, developmental course, and prognosis. Across cultures, taxonomic efforts are important for determining whether similar patterns of problems occur in different cultures and whether differences in patterns and prevalence provide clues to etiological factors that may be associated with culture.

The empirically based approach views assessment and taxonomy as two facets of the process of identifying important similarities and differences between individuals. According to this view, *assessment* is the identification of the distinguishing features of each individual. *Taxonomy* is the grouping of individuals according to their distinguishing features.

Our assessment procedures obtain data that are then analyzed to identify patterns of problems that tend to co-occur. Patterns of problems that are statistically robust across multiple analyses provide a basis for constructing syndrome scales on which to score individuals. That is, an individual's scores on the items of a syndrome are summed to obtain a total score for that syndrome. The individual's syndrome score can then be compared to the distributions of scores obtained on that syndrome by large samples of peers.

By displaying an individual's score on a profile comprising all the syndromes relevant to the individual's age and sex, practitioners can easily see whether the problems reported for the individual deviate much from the problems reported for peers. If the individual's scores are deviant, the profile of syndrome scores clearly indicates the areas in which the individual is deviant. The taxonomy of problem patterns derived from stand-

ardized assessment of large samples of individuals is thus, in turn, applicable to the assessment of new individuals. The empirically derived syndrome scales aggregate numerous specific problems into a relatively small number of taxonomic patterns for evaluating individuals, as well as for conducting research on new samples.

Cross-Cultural Applications

We have developed a family of standardized assessment instruments for obtaining data from parents, teachers, clinical interviews, trained observers, and the subjects themselves. From data on large clinical samples, we have derived eight cross-informant syndromes that can be scored from parent, teacher, and self ratings of both sexes between the ages of 4 and 18 (Achenbach, 1991 a). Because no one informant is likely to provide a complete picture of a youngster's functioning, a computer program is available for comparing data from different informants in terms of specific problems, empirically derived syndromes, and profile patterns identified via cluster analysis (Achenbach, 1993). Counterparts of some of the eight cross-informant syndromes can also be scored from related instruments for obtaining standardized assessment data from clinical interviews (McConaughy & Achenbach, 1994), direct observations (Achenbach, 1991b), and parent and self-reports for young adults (Achenbach et al., in press).

We will focus here on the assessment of children and adolescents via parent, teacher, and self-reports, which have stimulated the most numerous and rigorous cross-cultural applications of our empirically based approach. Translations of the assessment instruments are available in 40 languages, findings have been published from 28 cultures, and applications of the instruments have been reported in some 1,500 publications (Brown & Achenbach, 1994).

Strategies for Cross-Cultural Comparisons

An initial goal of most cross-cultural applications of the empirically based approach is to compare the distributions of problem scores obtained in samples that are representative of different cultures. By applying the same assessment procedures to samples from different cultures, we can determine whether a procedure developed in one culture

yields similar results in a different culture. A finding of similar results supports use of the same assessment procedures for coordinating research, clinical communication, and training between the two cultures. Multicultural work that can use the same procedures extends research opportunities beyond those that are typically available in a single culture. Furthermore, findings that replicate in multiple cultures inspire more confidence than those that are found in only one culture.

On the other hand, if an assessment procedure yields significantly different results for representative samples from different cultures, this opens the door to research on the possible reasons for the differences. The reasons might range from methodological factors — such as differential use of response scales — to differences in cooperation by informants, and to cultural, biological, and environmental factors that are associated with the different results. By pinpointing reasons for cross-cultural differences, we can enlarge our understanding of the factors that may affect both psychopathology and its assessment.

The following steps are usually required for rigorous cross-cultural comparisons:

1) Extensive research to develop instruments in Culture A, including thorough tests of reliability and validity.
2) Careful translation, back translation, and pilot testing of the instruments in the language of Culture B to capture the instruments' content in ways that are clear to the intended type of informant.
3) Application of the translated instruments to representative samples of people in Culture B.
4) Tests of the reliability and validity of the instruments in Culture B.
5) Rigorous statistical comparisons of the scores obtained by similar samples from Cultures A and B. Comparisons should employ samples that are similar with respect to variables that may affect scores within cultures, such as the age, sex, and referral status of the subjects, the type of informant, and socioeconomic status (SES).
6) To obtain a well-differentiated picture of similarities and differences between cultures, comparisons should be done at the relatively molecular level of specific items and at more molar levels, such as syndromes and other higher order groupings of items. Comparisons should also be made in relation to the types of

informants, plus subject characteristics, such as age, sex, referral status, and SES.
7) Patterns of co-occurring problems should be compared between the cultures, as can be done by confirmatory factor analysis.

Cross-Cultural Findings

Inclusion of Cross-Cultural Studies

The studies included in the present overview followed the procedures outlined above and used translations of the CBCL (Child Behavior Checklist, 1991 b), and/or TRF (Teacher's Report Form; Achenbach, 1991 c), and/or YSR (Youth Self-Report; Achenbach, 1991 d). We sought studies published in English that applied these instruments to samples representative of children and/or adolescents from a particular culture. However, we also included studies, such as a study of Chinese children (Weine et al., 1994), pertaining to samples that were not fully representative of the population, but that were rigorously matched with the comparison sample on age, sex, SES, and, in case of the Chinese study, degree of urbanization.

Results

Table 1 summarizes 14 studies of 10 cultures that were included in our overview. For brevity, we will reference these studies according to the numbers listed in Table 1.

Studies were excluded if they were not published in English (Montenegro, 1983; Lösel et al., 1989) had a response rate below 80% (Remschmidt & Walter, 1990), or used nonrandomized samples (Schwager et al., 1982).

All studies reported here employed rigorous statistical comparisons between scores from pairs of cultures. Seven studies employed subjects from a sample representative of the U.S.A. (Achenbach et al., 1991). Subjects from this sample were matched to subjects from other cultures on key demographic variables such as sex, age, and, if possible, SES and degree of urbanization. Five other studies drew subjects from a random sample of the Washington D.C. area of the U.S.A. (Achenbach & Edelbrock, 1981). Additional studies included one where the comparison subjects were Dutch (Hellinckx et al.,

Table 1. Cross-cultural comparison study using the CBCL, TRF and YSR.

Study	Nationality	Informant[a]	Age	N[b]	Sample	Comparison[c] sample	Reported level of analysis
1. Achenbach et al. (1989)	Puerto-Rico	P, T, S	4–16	724 CBCL 567 TRF 274 YSR	multistage probability throughout Puerto-Rico	US-East Coast for the CBCL, US-Nebraska, Tennessee and Pennsylvania for TRF (Achenbch & Edelbrock, 1986), and US-Worchester, Massachusetts for the YSR (Achenbach & Edelbrock, 1987) matched for SES	total problem score
2. Achenbach et al. (1990)	Australia-Sydney	P	4–16	1300 CBCL	multistage probability sample from Sydney	US-East Coast; N = 1300	items total problem score
3. Achenbach et al. (1987)	The Netherlands	P	4–16	2033 CBCL	two-stage probability sample from the province of Zuid-Holland	US-East Coast; N = 1300	items total problem score
4. Achenbach et al. (1987)	The Netherlands	T	6–11	748 TRF	two-stage probability sample from the province of Zuid-Holland	US-East Coast; N = 600	items total problem score
5. Hellinckx et al. (1994)	Belgium	P	6–12	1120 CBCL	two-stage probability sample from Flemish province	Dutch normative sample; N = 1126	items scales total problem score
6. Lambert et al. (1994a)	Jamaica	P	6–11	360 CBCL	two-stage probability sample from schools in Kingston and rural North-eastern Jamaica	US-National; N = 946	items scales total problem score
7. Lambert et al. (1994b)	Jamaica	T	6–11	359 TRF	two-stage probability sample from schools in Kingston and rural North-eastern Jamaica	US-National; N = 665	items scales total problem score
8. McDonald et al. (1994)	Greece-Athens	P	6–11	356 CBCL	random sample from schools selected to be geographically representative	US-National; N = 356 matched on age, sex, informant and SES	scales total problem score
9. Stanger et al. (1994)	France	P	6–11	921 CBCL	all children from 13 public and 5 private schools randomly chosen from the Eure-et-Loire department	US-National; N = 921 matched on sex, informant, and for the majority on age, and SES	items scales

Table 1. (continued)

Study	Nationality	Informant[a]	Age	N[b]	Sample	Comparison[c] sample	Reported level of analysis
10. Verhulst et al. (1993)	The Netherlands	S	11–18	803 YSR	6-year follow-up sample of original two-stage probability sample from the province of Zuid-Holland	US-National; matched on sex, age, and SES	items scales total problem score
11. Weine et al. (1994)	China	P,T	6–13	364 urban 105 rural CBCL and TRF	non-random sample of 6 schools from urban and rural areas in the Fujina province	US-National; matched on sex and age, and for the majority on SES N = 364 urban; 105 rural	total problem score
12. Weisz et al. (1993)	Kenya	P	11–15	90 CBCL	sample from 3 locations in the Embu district	US-National (90 Caucasian and 32 African-Americans), and Thai (N = 96) sample	total problem score
13. Weisz et al. (1987)	Thailand	P	6–11	360 CBCL	random sample of 1 child from each selected class from 38 schools from Bangkok and 4 major regions	US-East Coast; N = 600	items total problem score
14. Weisz et al. (1989)	Thailand	T	6–11	359 TRF	random sample of 1 child from each selected class from 38 schools from Bangkok and 4 major regions	US-Nebraska, Tennessee and Pennsylvania sample (Achenbach & Edelbrock, 1986); N = 586	items total problem score

a   P = parent; T = teacher; S = Self-Report
b   CBCL = Child Behavior Checklist; TRF = Teacher's Report Form; YSR = Youth Self-Report
c   US-East Coast sample, see Achenbach & Edelbrock (1981)
    US-National sample, see Achenbach et al. (1991)

1994), one where the comparison subjects were from Nebraska, Tennessee, and Pennsylvania U.S.A. (Achenbach et al., 1990 a), and two where the comparison subjects were from Worcester, Massachusetts, U.S.A. (Achenbach et al., 1990 b; Weisz et al., 1989).

In all studies, the mean problem scores from one culture were compared with those from the other by analyses of variance, with SES as a covariate or as a main effect. In this way, SES effects were partialled out to control for possible differences in SES distributions across nationalities, sex, and age. In 10 studies the CBCL was used, in five the TRF, and in two the YSR.

Nationality Differences in Total Problem Scores

Table 2 lists the mean total problem scores reported in each study, except the French study (9), for each instrument. The last two rows list the range of mean problem scores for the two comparison samples from the U.S.A. These scores differ somewhat across the studies, because the composition of the U.S. samples varied as a result of demographic matching to different samples from other cultures.

The second column for each instrument shows the percentage of variance accounted for by significant nationality differences.

CBCL. Across the different samples, the mean total problem score for the CBCL ranged from 20.0 (U.S.-East Coast sample) to 35.4 in the Greek sample (8). The largest difference between matched samples was for Puerto Rico (1) versus the mainland U.S.A., accounting for 13% of the variance, with higher scores for the Puerto Rican sample. Of the 11 cross-cultural comparisons, six showed no significant difference between cultures.

According to Cohen's (1988) criteria for effect sizes, two significant nationality differences were small, while three were medium. All significant differences involved lower scores for the U.S.A. sample.

Several studies have tested similarities between the rank ordering of item scores by computing Pearson correlations between mean item scores in pairs of samples. These correlations (with numbers in brackets referring to studies in Table 1) were: 0.79 (1); 0.89 (2); 0.84 (3); 0.79 (8); 0.83 (9), 0.66 for urban Chinese, (11), 0.44 for rural Chinese (11), and 0.72 (13). Despite nationality differences in mean total problem scores, the rank order of the item scores did not differ much between most nationalities.

TRF. The mean total problem scores for the TRF ranged from 17.6 for the Dutch sample (4) to 39.0 for the rural Chinese sample (11). Of the six comparisons, four showed significant differences indicating lower scores for the comparison sample,

Table 2. Mean total problem scores for different nationalities and results of cross-national comparison.

Nationality[a]	CBCL % Mean	Variance[b]	TRF % Mean	Variance[b]	YSR % Mean	Variance[b]
Australia[2]	31.6	11				
Belgium[5]	21.1	n.s.[d]				
China-urban[11]	27.6	n.s.	29.4	<1		
China-rural[11]	31.3	n.s.	39.0	n.s.		
France[9]	n.g.[c]	3				
Greece[8]	35.4	8				
Jamaica[6,7]	23.2	n.s.	27.2	2		
Kenya[12]	26.6	n.s.[e]				
Te Netherlands[3,4,10]	20.5	n.s.	17.6	n.s.	24.0	13
Puerto Rico[1]	34.9	13	30.1	3	35.4	4
Thailand[13,14]	24.2	1	30.9	5		
USA-East Coast	20.0–20.8[f]	–	19.3–20.1[f]	–	43.9	–
USA-National	24.2–28.1[f]	–	20.7–29.4[f]	–	37.9	–

Note: [a] numbers in superscript refer to studies listed in Table 1; [b] percentage of variance accounted for by significant differences in mean total problem scores for each nationality versus U.S.A. samples, except for Belgium[5] (compared with Dutch sample); [c] n.g. = not given; [d] n.s. = not significant; [e] compared with subsample of Afro-Americans from U.S. national sample; [f] for U.S.A. comparison samples, the range of mean total problem score is given; means differ slightly as a result of matching the U.S.A. sample on the sample from the other nationality.

with three small effects according to Cohen's (1988) criteria, and one accounting for <1% of the variance.

YSR. The two cross-cultural comparisons of YSR scores revealed significantly higher scores in U.S.A. mainland samples than in the samples from Puerto Rico and the Netherlands (1,10). The self-report findings in the Puerto Rico sample contrasted with those for the parent and teacher reports, as parents and teachers scored Puerto Rican children significantly higher than mainland

Table 3. CBCL items showing the greatest and least similarity in comparisons between the U.S.A. and Australia, France, Netherlands, Jamaica and Thailand, and between The Netherlands and Belgium.

No significant differences in at least 5 of the 6 comparisons
6.	Bowel movements outside toilet
15.	Cruel to animals
16.	Cruelty, bullying, or meanness to others
18.	Deliberately harms self, or attempts suicide
20.	Destroys his/her own things
21.	Destroys things belonging to his/her family or others
24.	Doesn't eat well
28.	Eats or drinks things that are not food
30.	Fears going to school
48.	Not liked by other kids
49.	Constipated, doesn't move bowels
52.	Feels too guilty
53.	Overeating
56g	Vomiting, throwing up
66.	Repeats certain acts over and over; compulsions
67.	Runs away from home
70.	Sees things that aren't there
73.	Sexual problems
78.	Smears or plays with bowel movements
79.	Speech problems
82.	Steals outside the home
84.	Strange behavior
91.	Talks about killing self
105.	Uses alcohol or drugs for nonmedical purposes
106.	Vandalism
107.	Wets self during the day

Significant differences in at least 5 of the 6 comparisons
43.	Lying or cheating
45.	Nervous, highstrung, or tense
49.	Demands a lot of attention
56f	Stomachaches or cramps
58.	Picks nose, skin, or other parts of body
68.	Screams a lot
71.	Self-conscious or easily embarrassed
74.	Showing off or clowning
87.	Sudden changes in mood or feelings
92.	Talks or walks in sleep
98.	Thumb-sucking
102.	Underactive, slow moving, or lacks energy
104.	Unusually loud

Note: Comparisons refer to the following studies listed in Table 1: 2, 3, 5, 6, 9, 13.

children. Also, the much lower scores for Dutch versus U.S. self-reports contrasted with the lack of difference in mean scores for Dutch versus American parents' and teachers' reports.

Nationality Differences in Item Scores

To identify items that showed the greatest cross-cultural similarities and differences, we examined comparisons between item scores derived from pairs of cultures. Because U.S. YSR scores were compared with those from only two other cultures, we investigated nationality differences per item only for the CBCL and TRF.

Table 3 lists the CBCL items that showed no significant differences in at least five of the six comparisons, as well as items that did show significant differences in at least five of the six comparisons. These items can be regarded as the least versus the most sensitive to cultural factors. Similarly, Table 4 reports the most and least similar items across cultures for the TRF.

Table 4. TRF items showing the greatest and the least similarity in comparisons between the U.S.A. and Jamaica, The Netherlands and Thailand.

No significant differences in all 3 comparisons
16.	Cruelty, bullying, or meanness to others
18.	Deliberately harms self or attempts suicide
19.	Demands a lot of attention
25.	Doesn't get along with other pupils
35.	Feels worthless or inferior
55.	Overweight
56c	Nausea, feels sick
56f	Stomachaches or cramps
56h	Vomiting, throwing up
70.	Sees things that aren't there
82.	Steals
83.	Stores up things he/she doesn't need
96.	Seems preoccupied with sex
106.	Overly anxious to please
107.	Dislikes school

Significant differences in all 3 comparisons
2.	Hums or makes other odd noises in class
4.	Fails to finish things he/she starts
15.	Fidgets
61.	Poor school work
66.	Repeats certain acts over and over; compulsions
69.	Secretive, keeps things to self
86.	Stubborn, sullen, or irritable
88.	Sulks a lot

Note: Comparisons refer to the following studies from Table 1: 4, 6, 14.

Table 5. CBCL items and syndromes and TRF items showing sex differences.

CBCL	TRF
[1]CBCL items with sex differences in at least 5 of the 6 comparisons	[2]TRF items with sex differences in all 3 comparisons
7. Bragging, boasting	2. Hums or makes other odd noises in class
8. Can't concentrate, can't pay attention for long	4. Fails to finish things he/she starts
10. Can't sit still, restless, or hyperactive	10. Can't sit still, restless
23. Disobedient at school	15. Fidgets
37. Gets in many fights	37. Gets in many fights
41. Impulsive or acts without thinking	41. Impulsive or acts without thinking
61. Poor school work	61. Poor school work
72. Sets fires	62. Poorly coordinated or clumsy
74. Showing off, or clowing	78. Inattentive, easily distracted
94. Teases a lot	90. Swearing or obscene language
95. Temper tantrums or hot temper	
98. Thumb-sucking[G]	
[3]CBCL scales with sex differences in at least 5 of the 6 comparisons	
Somatic Complaints[G] Attention Problems	

Note: [G] girls were scored higher than boys; all other differences indicate higher scores for boys over girls; [1] comparisons reporting CBCL items refer to the following studies from Table 1: 2, 3, 5, 6, 9, 13; [2] comparisons reporting TRF items refer to the following studies from Table 1: 4, 7, 14; [3] comparisons reporting CBCL syndrome scales refer to the following studies from Table 1: 5, 6, 8, 9, 11 urban, 11 rural.

Sex Differences

Sex differences that are robust across different cultures may reflect factors that transcend specific cultures, such as biological factors, or cross-cultural similarities in socialization practices. To identify such differences, we looked for significant sex effects that were consistent across different cultures.

CBCL. Of the ten cross-cultural comparisons that tested sex differences in CBCL total problem scores, two showed significant sex differences (3,5). Both differences reflected higher scores for boys than girls, but accounted for less than 1% of the variance.

In contrast to the minimal sex differences in total problem scores, there were numerous sex differences in individual item scores and syndrome scale scores. Among the six cross-cultural comparisons that reported item analyses, we looked for items that showed a significant sex effect in at least five comparisons. Table 5 lists the 12 CBCL items that showed consistent sex differences. On 11 items, boys were scored higher, while girls were scored higher only on item *98. Thumb sucking.* Of the 11 items scored higher for boys, seven belonged to the higher order grouping of Externalizing problems, whereas the remaining four belonged to the Attention Problems syndrome, which is considered to be separate from both the

Externalizing and Internalizing groupings (Achenbach, 1991 a).

The same procedure was followed for CBCL syndrome scale scores. Cross-culturally consistent sex differences were found for the Somatic Complaints syndrome (higher scores for girls) and the Attention Problems syndrome (higher scores for boys).

TRF. In four of the five comparisons for four different nationalities (Chinese rural and urban samples were analyzed separately), teachers scored boys higher than girls. Three sex effects were small according to Cohen's (1988) criteria, while one accounted for less than 1% of the variance. Table 5 lists items that showed significant sex differences in all three cross-cultural comparisons of TRF items.

Cross-Informant Consistency in Sex Differences Across Cultures

On four items, both parents and teachers scored boys significantly higher than girls across cultures. These items were: *10. Can't sit still, restless, or hyperactive; 37. Gets in many fights; 41. Impulsive or acts without thinking,* and *61. Poor school work.* Furthermore, the CBCL item *8. Can't concentrate, can't pay attention for long,* on which parents scored boys

higher than girls across cultures, resembles TRF item 78. *Inattentive, easily distracted*, on which teachers scored boys higher than girls.

SES Differences

Cross-cultural studies that assessed SES in similar ways can be used to determine which SES differences were the least culture sensitive.

CBCL. Five of the six cross-cultural comparisons that tested SES effects on total problem scores reported significantly higher total problem scores for lower SES children. However, the SES effects were small in three studies, and accounted for less than 1% of the variance in two studies.

Table 6 lists the items and syndrome scales that showed cross-cultural consistency in SES effects. All SES effects indicated higher problem scores for lower SES children.

TRF. The two studies that assessed SES effects on TRF problem scores (4,7) reported significantly higher total problem scores for lower SES children, but both effects were small. Table 6 lists the

TRF items that showed significant SES differences in both comparisons.

Cross-Informant Consistency in SES Differences Across Cultures

As can be seen from Table 6, parents and teachers in multiple cultures consistently scored the following items higher for low SES children: *10. Can't sit still, restless, or hyperactive; 37. Gets in many fights*, and *93. Talks too much*.

Age Differences

Age had a significant linear effect on CBCL total problem scores in three out of nine comparisons (1,3,5). All three significant effects indicated higher scores for younger children, with two effects being small, and one effect accounting for less than 1% of the variance. For the TRF, only one significant age effect on total problem score was found in five comparisons (4,7,11,14). This effect indicated higher scores for older children (1% of variance).

On the level of CBCL items and scales, no non-chance age effects were found in the Greek and French studies (8,9). Of the remaining five com-

Table 6. CBCL items and syndromes and TRF items showing SES differences.

CBCL	TRF
[1]*CBCL items with SES differences in the same direction in at least 4 of the 5 comparisons*	[2]*TRF items with SES differences in both comparisons*
	2. Hums or makes other odd noises in class
8. Can't concentrate, can't pay attention for long	4. Fails to finish things he/she starts
10. Can't sit still, restless, or hyperactive	10. Can't sit still, restless, or hyperactive
11. Clings to adults or too dependent	15. Fidgets
37. Gets in many fights	37. Gets in many fights
38. Gets teased a lot	59. Sleeps in class
43. Lying or cheating	61. Poor school work
53. Overeating	62. Poorly coordinated or clumsy
90. Swearing or obscene language	78. Inattentive, or easily distracted
93. Talks too much	93. Talks too much
99. Too concerned with neatness or cleanliness	104. Unusually loud
[3]*CBCL scales with SES effects in at least 3 of the 4 comparisons* Social Problems Attention Problems Delinquent Behavior Externalizing	

Note: All SES effects indicated higher scores for lower SES children.
[1]comparisons reporting SES effects on CBCL items refer to the following studies from Table 1: 2, 3, 5, 6, 9; [2]comparisons reporting SES effects on TRF items refer to the following studies from Table 1: 4, 7; [3]comparisons reporting SES effects on CBCL syndrome scales refer to the following studies from Table 1: 5, 6, 8, 9.

parisons, only one item (*61. Poor schoolwork*) showed a significant age effect in all five comparisons, with higher scores for older children, whereas in four comparisons significant age effects were found for seven items, all indicating higher scores for younger children (*14. Cries a lot; 19. Demands a lot of attention; 20. Destroys his/her own things; 24. Does not eat well; 29. Fears certain animals, situations, or places, other than school; 108 Wets the bed;* and *109. Whining*). No consistent age effects were found on the CBCL syndrome scales.

For the TRF, no consistent age effects were detected in the three comparisons reporting age effects on item scores and/or syndrome scale scores.

Cross-Informant Correlations

Several studies reported correlations between ratings of the same subjects by different informants. We could therefore assess the typical level of cross-informant agreement using similar assessment methodologies in different cultures. As Table 7 shows, cross-informant correlations did not differ much from culture to culture. No cross-informant correlation from any study differed significantly from the mean correlation shown in Table 7 for each pair of informants, according to Fisher's *Z* test.

Cross-Cultural Tests of Syndromes

We tested whether the 1991 cross-informant syndromes for the CBCL, TRF, and YSR (Achenbach, 1991 a) would replicate in CBCLs for 4,674 Dutch children (De Groot et al., 1994). First, we determined the factor structure for the Dutch CBCL by performing exploratory factor analyses on the CBCL scores of 2,339 children

Table 7. Cross-informant correlations for different nationalities.

	Parent × Teacher	Parent × Self	Teacher × Self
China[1]	.31		
Jamaica[2]	.33		
Netherlands[3,4]	.32	.54	
Puerto Rico[5]	.35	.49	.25
U.S.A.[6]	.44	.41	.31
mean correlation	.35	.47	.28

Source: [1] Weine et al. (1994); [2] Lambert et al.(1994 b); [3] Verhulst & Akkerhuis (1989); [4] Verhulst & Van der Ende (1992); [5] Achenbach et al. (1990 a); [6] Achenbach (1991 a).

constituting a derivation sample. These analyses yielded syndromes very similar to the American cross-informant syndromes, except for the Social Problems syndrome. We then scored the Dutch children on both the American and Dutch versions of each syndrome. Next, we computed correlations between the syndrome scores obtained when the Dutch children were scored on the American version of the syndrome and when they were scored on the Dutch version.

As Table 8 shows, the cross-national correlations between the Dutch and the American scales ranged from 0.82 for the Social Problems syndrome to 0.99 for the Somatic Complaints and Anxious/Depressed syndromes. Table 8 also shows the proportion of items of the scale of one country not appearing on the scale of the other country.

We used confirmatory factor analysis to test the degree to which the Dutch exploratory syndromes and the American cross-informant syndromes fitted the scores in the cross-validation sample of 2,335 Dutch children. The Dutch exploratory syndromes and the American cross-informant syndromes both fit the cross-validation syndromes equally well according to both the Goodness of Fit Index, which was 0.885, and the Adjusted Good-

Table 8. Comparison of American and Dutch CBCL cross-informant scales.

Syndrome	r between American and Dutch scale scores	Proportion of items of American scale not on Dutch scale	Proportion of items of Dutch scale not on American scale
Withdrawn	0.94	2/9	1/8
Somatic complaints	0.99	0/9	1/10
Anxious/Depressed	0.99	1/14	1/14
Social Problems	0.82	4/8	3/7
Thought Problems	0.89	0/7	4/11
Attention Problems	0.94	3/11	3/11
Delinquent Behavior	0.94	2/13	3/14
Aggressive Behavior	0.98	5/20	4/19

ness of Fit Index, which was 0.878. The results thus showed great similarity between the Dutch and American syndrome structures.

The similarities between the Dutch and American CBCL syndromes, despite differences in language, culture, and mental health systems, supported the generalizability of the CBCL 1991 cross-informant syndromes across both countries. Similar analyses of Dutch and American TRF and YSR scores have also supported the cross-cultural generalizability of the syndromes (De Groot et al., 1995 submitted).

Summary and Conclusions

The empirically based approach uses standardized assessment procedures to score the behavioral and emotional problems of large samples of subjects. Syndromes of co-occurring problems are then derived from multivariate statistical analyses of problems reported for clinically referred subjects. The scores obtained by an individual can be used to determine how closely that individual's problems resemble each of the empirically derived syndromes. Scores on each syndrome and profile pattern of syndrome scores can be used clinically for targeting interventions and for evaluating changes in response to interventions.

As an example, if a clinician finds that a child is deviant on only the Attention Problems syndrome and meets diagnostic criteria only for attention deficit disorder, then attention problems would be the main focus of treatment. On the other hand, if the clinician finds that the child is deviant not only on the Attention Problems syndrome but also on other syndromes, such as the Anxious/Depressed or Aggressive Behavior syndrome, this would argue for interventions to deal with these additional problems, as well as the attention problems. Furthermore, to fully evaluate the child's response to interventions, the clinician needs to reassess not only the problems targeted for treatment, but also other problem areas as well. If we reassess only the problems that were treated, we might miss the fact that improvement in the target area was accompanied by lack of improvement or even worsening in areas reflected by other syndrome scales. By initially assessing children in terms of a profile of syndrome scales and then reassessing them later in terms of the same profile, the clinician can pinpoint areas of improvement, worsening, and no change.

Empirically based assessment instruments offer workers in different countries a common data language for describing and analyzing problems reported by various informants. Use of the same standardized assessment procedures in different countries makes it possible to identify cross-cultural variations in behavioral and emotional problems that may point to etiological factors. For example, comparisons of problem scores for representative samples of American versus Dutch children at two points in time have shown small but significant increases among American children that were not found among Dutch children (Verhulst et al., 1994). The increases occurred across a broad range of problems reported by American parents and teachers for children of both sexes, all SES levels, and different ethnic groups (Achenbach & Howell, 1993). These findings suggest that social conditions changed for many American children in ways that contributed to behavioral and emotional problems.

Use of the same standardized procedures also enables workers in one country to profit from findings by workers in other countries. Considering the paucity of researchers who specialize in child psychopathology, an additional advantage of standardized empirically based assessment is that it can increase the critical mass of workers focusing on similar problems by providing common tools with which to calibrate their efforts.

Progress in cross-cultural calibration is indicated by translations of the empirically based assessment instruments into 40 languages and by publications of findings from 28 cultures to date. We reviewed findings from rigorous comparisons of parents' reports of children's problems in various pairings of 11 cultures, teachers' reports in pairings of 6 cultures, and self-reports in pairings of 3 cultures. Significant differences were found between some pairs of cultures, but the cross-cultural differences in total problem scores were generally small. We also identified specific items that showed the fewest versus most cross-cultural differences, plus items and syndromes that consistently showed similar associations with sex and SES in most cultures that were compared. We found that the correlations between problem scores obtained from particular pairs of informants (parent × teacher; parent × self; teacher × self) were quite similar in multiple cultures. In addition, the cross-informant syndrome structure derived from the combination of parent, teacher, and self ratings of clinically referred American children was replicated for clinically referred Dutch children.

The empirically based approach to assessment and taxonomy provides concepts, methods, and tools that can foster international collaboration in multiple ways. This approach is compatible with diagnostic systems such as the ICD and DSM, as demonstrated by significant associations between diagnoses and empirically based syndromes in many studies (e.g., Weinstein et al., 1990; Gould et al., 1993; Chen et al., 1995). However, rather than defining disorders according to the same fixed cutpoints for all groups, the empirically based approach focuses on the actual distributions of problem scores obtained for relevant normative and clinical samples in each country. This means that variations in problems in relation to culture, as well as to such variables as sex, age, SES, and type of informant, can be readily detected and can be studied in their own right. Systematic documentation of such variations can contribute to improving the ICD/DSM approach as well as the empirically based approach and to promoting a more effective synthesis between them.

Acknowledgements

Résumé

Cet article rapporte une vue d'ensemble d'une évaluation fondée empiriquement avec taxonomie illustré par une recherche transculturelle en psychopathologie. L'approche empiriquement fondée utilise des procédures d'évaluation standardisées pour coter les problèmes comportementaux et affectifs dont dérivent les syndromes par analyse multi-variée. Les items et les syndromes sont cotés quantitativement pour refléter leur degré d'expression chez les individus, tels qu'ils sont rapportés par des informateurs spécifiques. Quoique l'approche pour évaluer les problèmes et construire des groupes taxonomiques diffère de cel de ICD/DSM, il n'y a pas de contradictions inhérentes entre leur modèle de troubles ni les traits des critères utilisés pour définir ces troubles. Des comparaisons transculturelles ont montré des différences relativement faibles dans les pourcentages de problèmes et la structure des syn-dromes, de plus une similarité considérable dans les associations des problèmes avec le sex et statut socio-économique, ainsi que des corrélations semblables entre les rapports par les différents types d'informateurs. La recherche sur les variations dans les problèmes en relation à la culture, le sexe, l'âge, le statut socio-économique et le type d'informateur peut contribuer à améliorer à la fois l'ICD/DSM et les approches empiriquement fondées et permettre une synthèse plus efficace entre eux.

Zusammenfassung

Diese Arbeit bietet eine Übersicht zur empirisch fundierten Erfassung und Taxonomie, dargestellt an der transkulturellen Psychopathologieforschung. Die empirisch fundierte Vorgehensweise beinhaltet die Verwendung von standardisierten Erhebungsverfahren zur Wertung von Verhaltens- und emotionalen Auffälligkeiten, aus denen Syndrome durch multivariate Analysen abgeleitet werden. Die Items und Syndrome werden quantitativ gewertet, um bei Einzelpersonen das Ausmaß der manifestierten Auffälligkeiten angeben zu können, die wiederum von bestimmten Informanten berichtet werden. Obwohl die Vorgehensweise zur Erfassung von Problemen und zur Bildung von taxonomischen Gruppen sich von dem ICD/DSM-Ansatz unterscheidet, gibt es keine inhärenten Widersprüche zwischen ihren Modellen für Störungen noch in kritischen Merkmalen, die zur Definition von Störungen verwandt werden. Transkulturelle Vergleiche haben relativ kleine Unterschiede in Problemraten und Syndromstruktur ergeben. Außerdem findet sich eine beträchtliche Ähnlichkeit im Hinblick auf die Assoziationen von Problemen mit Geschlecht und sozioökonomischem Status, aber auch ähnliche Korrelationen zwischen Berichten von unterschiedlichen Informanten. Forschung zur Variation von Problemen im Hinblick auf Kultur, Geschlecht, Alter, sozioökonomischen Status und Typus des Informanten können dazu beitragen, sowohl das ICD/DSM-Modell als auch andere empirisch fundierte Vorgehensweisen zu verbessern und eine effektivere Synthese zwischen diesen Vorgehensweisen zu erreichen.

References

Achenbach, T. M. (1966). The classification of children's psychiatric symptoms: A factor-analytic study. *Psychological Monographs, 80.*

Achenbach, T. M. (1991 a). *Integrative Guide for the 1991 CBCL/418, YSR, and TRF Profiles.* Burlington, VT: University of Vermont Department of Psychiatry.

Achenbach, T. M. (1991 b). *Manual for the Child Behavior Checklist/418 and 1991 Profile.* Burlington, VT: University of Vermont Department of Psychiatry.

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

Achenbach, T. M. (1991 d). *Manual for the Youth Self-Report and 1991 Profile.* Burlington, VT: University of Vermont Department of Psychiatry.

Achenbach, T. M. (1993). *Empirically Based Taxonomy: How to Use Syndromes and Profile Types Derived from the CBCL/418, TRF, and YSR.* Burlington, VT: University of Vermont Department of Psychiatry.

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 children aged four through sixteen. *Monographs of the Society of Research in Child Development, 46.*

Achenbach, T. M. & Edelbrock, C. (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. & Edelbrock, C. (1987). *Manual for the Youth Self-Report and Profile.* Burlington, VT: University of Vermont Department of Psychiatry.

Achenbach, T. M., Verhulst, F. C., Baron, D. G. & Akkerhuis, G. W. (1987 a). 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 & Adolescent Psychiatry, 26,* 317–325.

Achenbach, T. M., Verhulst, F. C., Edelbrock, C. S., Baron, D. G. & Akkerhuis, G. W. (1987 b). Epidemiological comparisons of American and Dutch children: II. Behavioral/emotional problems and competencies reported by teachers for ages 6 to 11. *Journal of the American Academy of Child & Adolescent Psychiatry, 26,* 326–332.

Achenbach, T. M., Conners, C. K., Quay, H. C., Verhulst, F. C. & Howell, C. T. (1989). Replication of empirically derived syndromes as a basis for taxonomy of child/adolescent psychopathology. *Journal of Abnormal Child Psychology, 17,* 299–323.

Achenbach, T. M., Bird, H. R., Canino, G. Phares, V.,

Gould, M. S. & Rubio-Stipec, M. (1990 a). Epidemiological comparisons of Puerto Rican and U.S. Mainland children: Parent, teacher, and self-reports. *Journal of the American Academy of Child & Adolescent Psychiatry, 29,* 84–93.

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

Achenbach, T. M., Howell, C. T., Quay, H. C. & Conners, C. K. (1991). National survey of problems and competencies among four- to sixteen-year-olds: Parents' reports for normative and clinical samples. *Monographs of the Society for Research in Child Development, 56.*

Achenbach, T. M., Howell, C. T. (1993). Are the American Children's problems getting worse? A 13-year comparison. *Journal of the American Academy of Child & Adolescent Psychiatry, 32,* 1145–1154.

Achenbach, T. M., Howell, C. T., McConaughy, S. H. & Stanger, C. (1995). Six-Year Predictors of Problems in a National Sample: 3. Transitions to Young Adult Syndromes. *Journal of the American Academy of Child & Adolescent Psychiatry,* in press.

American Psychiatric Association. (1980, 1987, 1994). *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed., 3rd ed., revised, 4th ed.). Washington, DC: Author.

Brown, J. S. & Achenbach, T. M. (1994). *Bibliography of Published Studies Using the Child Behavior Checklist and Related Materials: 1994 Edition.* Burlington, VT: University of Vermont Department of Psychiatry.

Chen, W. J., Faraone, S. V., Biederman, J., & Tsuang, M. T. (1995). Diagnostic accuracy of the Child Behavior Checklist scales for Attention-Deficit Hypoactivity Disorder: A receiver-operating characteristic analysis. *Journal of Consulting and Clinical Psychology, 62,* 1027–1025.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

De Groot, A., Koot, H. M. & Verhulst, F. C. (1994). The cross-cultural generalizability of the CBCL cross-informant syndromes. *Psychological Assessment, 6,* 225–230.

De Groot, A., Koot, H. M. & Verhulst, F. C. (1995). *The Cross-Cultural Generalizability of the TRF and YSR Cross-Informant Syndromes.* Submitted for publication.

Edelbrock, C. S. & Achenbach, T. M. (1980). A typology of child behavior profile patterns: Distribution and correlates for disturbed children aged 6–16. *Journal of Abnormal Child Psychology, 8,* 441–470.

Edelbrock, C. S. & Costello, A. J. (1988). Convergence between statistically derived behavior problem syndromes and child psychiatric diagnoses. *Journal of Abnormal Child Psychology, 16,* 219–231.

Ferdinand, R. F., Verhulst, F. C. & Witnitzer, M. (1995). Continuity and Change of Self-Reported Problem Behaviors from Adolescence into Young Adulthood. Journal of the American Academy of Child & Adolescent Psychiatry, in press.

Gould, M. S., Bird, H. & Jaramillo, B. S. (1993). Correspondence between statistically derived behavior problem syndromes and child psychiatric diagnoses in a community sample. Journal of Abnormal Child Psychology, 21, 287–313.

Hellinckx, W., Grietens, H. & Verhulst, F. C. (1994). Competence and behavioral problems in 6- to 12-year old children in Flanders (Belgium) and Holland: A cross-national comparison. Journal of Emotional and Behavioral Disorders, 2, 130–142.

Kraepelin, E. (1883). Compendium der Psychiatrie. Leipzig: Abel.

Lambert, M. C., Knight, F., Taylor, R. & Achenbach, T. M. (1994 a). Epidemiology of behavioral and emotional problems among children of Jamaica and the United States: Parent reports for ages 6 to 11. Journal of Abnormal Child Psychology, 22, 113–128.

Lambert, M. C., Knight, F., Taylor, R. & Achenbach, T. M. (1994 b). Comparisons of behavioral and emotional problems among children of Jamaica and the United States: Teacher reports for ages 6–11. Journal of Cross-Cultural Psychology, in press.

Lösel, F., Bliesener, T. & Köferl, P. (1989). Erlebens- und Verhaltensproblemen bei Jugendlichen: Deutsche Adaption und kulturvergleichende Überprüfung der Youth Self Report Form der Child Behavior Checklist. Nürnberg, Germany: Universität Erlangen Institut für Psychologie.

McConaughy, S. H. & Achenbach, T. M. (1994). Manual for the Semistructured Clinical Interview for Children and Adolescents. Burlington, VT: University of Vermont Department of Psychiatry.

McDonald, V. M., Tsiantis, J., Achenbach, T. M., Stefanidi, F. M. & Richardson, S. C. (1994). Competencies and problems reported by parents of Greek and American children ages 6–11. European Child & Adolescent Psychiatry, 4, 1–13.

Montenegro, H. (1983). Estandardacion de Inventorio de Problemas Conductales y Destrezas Sociales de T Achenbach en Ninos de 6 a 11 Anos. Santiago, Chile: Centro de Estudios de Desarrollo Y Estimulacion Psicosocial.

Quay, H. C. (1979). Classification. In H. C. Quay & J. S. Werry (Eds.), Psychopathological Disorders of Childhood (2nd ed.). Wiley: New York.

Remschmidt, H. & Walter, R. (1990). Psychische Auffälligkeiten bei Schulkindern. Zeitschrift der Kinder- und Jugendpsychiatrie, 18, 121–132.

Rey, J. M. & MorrisYates, A. (1992). Diagnostic accuracy in adolescents of several depression rating scales extracted from a general purpose behavior checklist. Journal of Affective Disorders, 26, 7–16.

Schwager, W., Dennie, D. & Smith, M. (1982). Franco-Ontarian Children in Sudbury. Sudbury Ontario: Laurentian University Department of Sociology and Anthropology.

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

Verhulst, F. C. & Akkerhuis, G. W. (1989). Agreement between parents' and teachers' ratings of behavioral/emotional problems of children aged 4–12. Journal of Child Psychology and Psychiatry, 30, 123–136.

Verhulst, F. C. & Van der Ende, J. (1992). Agreement between parents' reports and adolescents' self-reports of problem behavior. Journal of Child Psychology and Psychiatry, 33, 1011–1023.

Verhulst, F. C., Achenbach, T. M., Ferdinand, R. F. & Kasius, M. C. (1993). Epidemiological comparisons of American and Dutch adolescents' self-reports. Journal of the American Academy of Child and Adolescent Psychiatry, 32, 1135–1144.

Verhulst, F. C., Achenbach, T. M., Van der Ende, J. & Rietbergen, G.A.P. (1995). Long-term Changes in Problems Reported by Parents and Teachers of American Versus Dutch Children. In preparation.

Weine, A. M., Phillips, J. S. & Achenbach, T. M. (1994). Behavioral and emotional problems in Chinese and American children: Parent and teacher reports for ages 6–13. Journal of Abnormal Child Psychology, in press.

Weinstein, S. R., Noam, G. G., Grimes, K., Stone, K. & Schwab-Stone, M. (1990). Convergence of DSM-III diagnoses and self-reported symptoms in child and adolescent inpatients. Journal of the American Academy of Child & Adolescent Psychiatry 29, 627–634.

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

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

Weisz, J. R., 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.

World Health Organization. (1992). Mental Disorders: Glossary and Guide to Their Classification in Accordance with the Tenth Revision of the International Classification Diseases. Geneva: Author.