Acta Psychiatr Scand 1999: 100: 2–52
Printed in UK. All rights reserved

ACTA PSYCHIATRICA
SCANDINAVICA
ISSN 0902-4441

The Child Behavior Checklist (CBCL) and related material: standardization and validation in Danish population based and clinically based samples

N. Bilenberg

Bilenberg N. The Child Behavior Checklist (CBCL) and related material: standardization and validation in Danish population based and clinically based samples.
Acta Psychiatrica Scand 1999: 100: 2–52. © Munksgaard 1999.

The Child Behavior Checklist (CBCL) and related material, developed by Achenbach and Edelbrock in Vermont, was validated in a mailed survey. A population based sample of 779 children between the age of four and 17 years was compared to a sample of 146 children referred for child psychiatric service. Danish children scored very much like our Scandinavian and German neighbours, but low compared to most others. The CBCL mean 'total behavior problem score' in the population was 17.7. The checklists, especially the parent and teacher versions, provided good construct validity. Youths generally reported more emotional problem behavior than their parents and teachers did about them. In general, parents and youths agree more, reporting emotional problems, and parents and teachers agree more, when scattering externalizing behavior. Short screening constructs are introduced, and by the use of latent trait analysis, four clinically relevant sub-scales were generated. Predictive value, sensitivity, specificity and clinical validity must be undertaken in a future two-phase study.

Key words: Child Behavior Checklist (CBCL); epidemiology; latent trait analysis; standardization; validity

Niels Bilenberg, Institut for Psykiatrisk Grundforskning, Afdeling for Psykiatrisk Demografi, Psykiatrisk Hospital i Århus, DK-8240 Risskov

INTRODUCTION

Standardization or validation of psychometric instruments has become an important task for researchers throughout the psychiatric field. This discipline has grown due to changes in the psychiatric diagnostic base. Etiologically based circumstances, which were formerly essential in the process of correct diagnostic labeling, have now been replaced by well defined descriptive and operational criteria (4–6). Naturally, the collection of sufficient information to provide diagnostic reliance has also been modified. Numerous checklists, scales, structured interviews and other assessment instruments have been developed (7, 8). New psychometric instruments appear in the literature, although only a few of these remain to become a subject of international interest. Some research groups and authors do not hesitate to use an instrument without prior standardization in their context population. Due to the fact that competencies, behavior and emotions in children are dependent upon cultural context, this procedure should most certainly be avoided.

The aim of the study

The primary aim of this study is to introduce a validated screening and psychometric instrument within Danish Child and Adolescent Psychiatry for

Abbreviations:
ADHD: attention deficit hyperactive disorder; ANOVA: analysis of variance; ANCOVA: analysis of covariance; CBCL: child behavior checklist; CII: cross informant items; CIS: cross informant syndromes; DSM-IV: Diagnostic and Statistical Manual (of mental disorders), 4th edition; ICD-10: International Classification of Diseases, 10th edition; OCD: obsessive compulsive disorder; ROC: receiver operating characteristic; s.d.: standard deviation; SES: socio economical status; TRF: Teachers Report Form; YSR: Youth Self Report.

clinical, as well as for research purposes. In selecting the Child Behavior Checklist (CBCL) and related material by Achenbach and Edelbrock, standards and possible interpretations would assist, also, those allied professionals within clinical Child Psychology and Pediatrics, who are familiar with the checklists. Secondly, it is the intention to present some standard outcomes, which may be compared and discussed in relation to outcome from similar studies in other cultures. Finally, a combination of CBCL items, which would assist in screening was, also, requested.

During the process, a number of challenges and pitfalls in relation to the instruments appeared; these will be discussed later. Furthermore, a statistical approach differing from previously used methodology in the available CBCL literature will be introduced.

A review of the literature

The origin of the CBCL

The CBCL and related material was developed by psychologist Thomas M. Achenbach and child psychiatrist Craig Edelbrock in Vermont, US (9). On a nationwide basis, the CBCL has now become one of the most frequently used instruments in the collection of data regarding child behavior in clinical settings and in the purpose of research. The CBCL, together with Teachers Report Form (TRF) and Youth Self Report (YSR), are now available in more than 50 languages and they are formally standardized in several countries and cultures around the world (10, 11). The three checklists were generated from a pool of competence and problem items. The process of refinement is described by the authors in their early papers, as well as in the manuals (9, 12–15). The manuals present the validation and standardization procedures made on larger US samples. The 1991 profile was based on a total of 4220 equally numbered, demographically matched referred and non-referred children. Validity, reliability, discrimination and factorial analysis are evaluated, and scoring relating to each instrument is demonstrated in detail. Hand scoring-profiles and computer software regarding the scoring of all related questionnaires are available from the University of Vermont.

The CBCL outside the US

Verhulst and his co-workers in Rotterdam, The Netherlands, were the first to introduce the CBCL to Europe in 1982 (1–3). They provided a well designed two-phased epidemiological study of Dutch children, which included a profound standardization of the CBCL and TRF. Together with a number of researchers from France, Greece, Australia, Puerto Rico and China, T.M.Achenbach has generated national norms and investigated cross cultural variation (16–20). Achenbach, Verhulst and co-workers in particular, have performed a number of detailed comparisons of American and Dutch children, based on output from the CBCL and related instruments (21–25). Major cross-cultural resemblances are generally accepted in symptom prevalence, as well as in syndrome structures, although some differences have occurred and these will be discussed later (10, 26–28).

Terminology in validation and standardization

Validation and standardization of measurement scales and psychometric instruments have been

Table 1. Some commonly used concepts in this thesis

Concept	Brief definition
Validity	The *validity* of a questionnaire means how well it measures what it claims to measure. Lack of validity is systematic error
Construct	A *construct* is a variable that is abstract and latent rather than concrete and observable. In questionnaires an item is the observable aiming of a latent variable
Construct validity	*Construct validity* is the precision by which the combination of items in a questionnaire reflects the underlying phenomenons of interest
Factor analysis	*Factor analysis* and *latent trait (i.e.latent structure) analysis* is part of the construct validity, measuring which of several observable items, that *correlate* to an underlying trait or structure. These can be isolated and their mutual relation can be analyzed
Item bias	*Item bias* reflects the fact that some items must be interpreted differently when applied to ex. different sex, age or SES groups. This is a matter of concern in every validation study
Content validity	*Content validity* of a questionnaire means that all relevant aspects of interest are included and the weighting of items reflects their importance. This part of the validation procedure is documented by demonstrating the strategies that are used throughout the process. Strategies often rely on clinical judgments, ex. when constructing models in regression analyses or defining inclusion and exclusion criteria in a factorial analysis, therefore one must make these decisions and the underlying values explicit
External validity	*External validity* is testing to which degree the instrument succeeds in predicting the actual criterion, given as a *golden standard.* This type of validity is also called *criterion validity*
Reliability	The *reliability* of a questionnaire is the extent to which the measurements are free from unsystematic or random errors. Reliability can also be defined as repeatability. *Test-retest reliability* is the repeatability over time, where the same informer fills in the questionnaire on two different occasions. *Inter-rater reliability* is repeatability across informers

established upon a set of conventionalized procedures. As a result of this a specific terminology has evolved (29, 30). Unfortunately, some terms are utilized differently by various authors, requiring precision in the currently used definitions. Some essential concepts and terms are stipulated in Table 1.

Normative sampling

Calibration of a metric instrument must be performed in a normative sample representative of the population in question. In extremely inhomogeneous multi-ethnic populations, each significant sub-group must be represented in numbers which are appropriate for individual statistical analysis. A variety of different sampling procedures have been used in standardization of the CBCL material. This, together with varying exclusion criteria influence the possible representation and relevance of comparisons between the analyzed normative samples. Telephone directory identification of families invited to participate was used for sampling in the US (9), selected school samples were used in The Netherlands (1), Sweden (31), France (16), Switzerland (32), Greece (17) and other countries. In Iceland (33) and Norway (34), as in this study, the population register was used to draw sex and age stratified samples.

As a consequence of exclusion from population based samples of children attending special classes, and children referred to child psychiatric or child psychologic services, various degrees of super-samples have emerged. Selection bias has also been induced by sampling from areas or populations not representative of whole nations. This was probably the case in the Swedish study, as well as in the first Greek study. Even the extensive US non-clinical sample, which supports the US profiles in the manuals, has been criticized for not being representative. Method of data collection, interviews instead of mailed survey, has also been considered as atypical use of material (35).

The number of subjects in normative samples vary dramatically, ranging from a study of 90 Kenyan children (36) to 2856 German children in a recent study (37). Varying age groups have been analyzed, all somewhere in the span between four and 18 years.

Administration

Administration of the questionnaires presented to parents, teachers and youths has been conducted in various ways. Parents in the US non-clinical sample and in the Dutch sample were interviewed by way of visits to the home. In other cases teachers have collected the completed CBCL forms from parents. Teachers and youths have normally answered the TRF and YSR by way of questionnaire. In some school-based studies, youths have completed the questionnaires in the classroom. Mailed surveys such as the present study have also been applied in Iceland, Norway and Germany.

Response rates and drop-out analysis

Response rates must be high to postulate complete representation of a normative sample. However, 100% participation is hardly ever possible and drop-out analysis can compensate by illuminating selection bias. Response rates vary from less than 50% to almost 100%. Drop-out analyses have only been performed in a few standardizations. In other studies, authors have attempted to illustrate the representation of responders by using different available demographic data as comparisons. The availability of demographic data regarding the general population, and register data relating to children and families, varies from country to country. Public registration to the degree found in Denmark is extremely rare. The Danish Statistical Institute is beneficial in attrition analysis, making it possible to conduct epidemiologic studies without ideal participation, whilst still producing valid results. The best design for drop-out analysis is actually obtaining the relevant information (in this case completed questionnaires) from a random sample of non-responders and comparing these to the responders in the original sample. This process has only been conducted in Germany (38, 39).

Statistical analysis

Statistical analysis tends to follow the same methodological procedures, initiated by Achenbach in his original work. Single item discrimination between population based and clinically based samples have been measured by the analysis of variance and co-variance (ANOVA and ANCOVA) in sex and age strata, and with socio-economical status (SES) as covariate. Cross-informant syndromes have been generated from the clinically based sample by standard factor analytic procedures summarized in the manuals (9).

Following standard procedures is logical, when comparisons are the aim of a study and authors wish to contribute to the literature regarding cross-cultural conditions. Nevertheless, it is appropriate to question the traditions and, perhaps, introduce new pathways for relevant interpretation of the CBCL material. In this study, another discriminant analysis is introduced, and standard factor analysis is supplied with latent trait analysis of dichotomized items. The US cross-informant syndromes do not fulfil criteria as scales in the psychometric sense of the term (i.e. higher scoring represents more severe pathology). Much too low factor loadings were accepted when tying items to the different components, and items were included without consideration of whether they reflected a continuum in severity.

The CBCL in epidemiological studies

The CBCL and related material have, in a number of studies, been used as screening or as the first step in two-phase designs (1, 2, 34, 40, 41). It has, also, been implemented as a predictor of outcome in longitudinal studies (42–49). As a screening instrument, the CBCL appears to be adequate and as good as the 'Rutter Parental Questionnaire' and 'Quay-Peterson: Revised Behavior Problem Checklist' for use in the general population (50). Predictive power, sensitivity, specificity and best cutoff of the CBCL materials in screening are in some studies evaluated by Receiver Operating Characteristics (ROC) analysis (34, 51). In this study, because of limitations in the used design, only sensitivity (correct classification of cases) will be tested and discussed.

Both externalizing and internalizing abnormal behavior in children has considerable long-term stability, although externalizing behavior predicts the worst form of clinical outcome (48). From follow-up studies of teacher-reported abnormal behavior, it appeared that disturbed girls, in particular, were stable and persistently deviant throughout a four year period (49). Children with high, or deviant score, on two or more sub-scales (comorbidity), had a poorer prognosis than children who presented high scoring on one sub-scale only (47).

The CBCL material has never provided good diagnostic validity in child and adolescent psychiatric clinic. Nevertheless, as a helpful instrument in visitation and as a guideline for early diagnostic categorizing, the material can be widely implemented in clinical settings. Some sub-scales of problem items, in particular the 'attention problems' and the 'anxious/depressed' factor have proven acceptable diagnostical accuracy (52–55).

The CBCL in pediatrics

The CBCL has made a significant impact in pediatric literature. The general aim of a number of studies have been to compare pediatric diagnostic sub-groups with control groups, or normative CBCL data taken from population based samples. Chronically-ill children (56), children with nephrotic syndrome (57), juvenile arthritis (58, 59), inflammatory bowel disease (60), congenital heart disease (61), Prader-Willi syndrome (62) and many others have been analyzed. The results have varied from no impact in behavioral status to significant differences between groups. The risk that matched controls are too well-functioning, favors comparisons to representative population based samples in this type of study (63).

Some obstacles concerning the CBCL

In a particular article, Drotar et al. (64) focused on a number of difficulties relating to actual interpretation of the CBCL material. They questioned the value of the checklists within the normal range. In addition, in a number of studies the CBCL and YSR have shown low discriminative power in the aspect of social competence. The possibility that physical symptoms (items 56 a–h) reflect an acute, or even chronic medical condition, instead of behavioral or psychological difficulties, was also stated. Finally, the use of norms, as opposed to comparison groups, was questioned (i.e. when normative samples are not tested for representation and owing to exclusion criteria, represent super-samples).

MATERIAL

The study population

The study is based on a population of children between the ages of four and 17 years, living in the Fynen area, one of Denmark's main islands. The island itself is situated in the center of Denmark and is surrounded by a number of smaller islands, all covering an area of approximately 3485 square kilometres. In 1996 approximately 470 000 inhabitants lived in the area. Of these inhabitants, approx. 250 000 live in the city of Odense and suburbs. In addition, there are four major towns boasting 10 000 to 30 000 inhabitants. The remainder live in smaller towns and rural districts. At 1st January 1996, a total of 67 384 children were included in the study population. The county of Fynen has been used in a number of epidemiologic studies, due to the fact that the population represents the total Danish population in almost all sociological and demographical aspects (65). The Child and Adolescent Psychiatric Department in Odense is the only one of its kind and receives all referrals for child psychiatric service within the county.

The normative sample

A stratified population based sample of 1300 children, 50 boys and 50 girls born each year between 1979 and 1991, was taken from the Danish public register. All children were born in early September which ensured that they would be midway between birthdays, at the time they received the questionnaire.

The clinically based sample

The research group was supplied with the names, addresses and birth codes of 326 children, who were referred for child psychiatric service within the county of Fynen between Autumn 1995 and 31 October 1997. Children more than 16 years of age, and those who had previously received outpatient treatment at the time of data collection, were excluded. During this two year period, 302 referred children and their families received the material by mail. Part of the actual design was that no connection between the CBCL project group and the clinical assessment should be made, thus making data collection in the clinically based sample and in the population based sample more comparable.

Data collection and data sources

The mailed survey

Questionnaires were sent to each individual proband in the population based sample during March 1996. The probands in the clinical sample were approached consecutively throughout the entire period. The envelope itself was addressed to the child himself, whilst the content of the envelope was directed at parents and child. All subjects received a white formula, containing instructions as to how the various forms should be completed, together with questions relating to family constellation, residence (urban or rural), parents' work and educational level, economic situation, whether the child had had contact with child psychiatric departments, or psychological services within the last year, as well as questions regarding the well-being of the family itself. Participants between the age of six and 17 years were asked to take the TRF to the classroom teacher. Prepaid envelopes were provided for the families, as well as the schools. The informants (parents, youths and teachers) were requested to complete the questionnaires without consulting each other, and teachers were asked to return the forms direct to the research group. In the case of non-repliers two mailed reminders were sent. Two weeks later, non-responders received a postcard and after a further two week period, the material was re-mailed to those who had not yet reacted. Exactly the same procedure was used in the clinical sample, except that data collection was spread over a one year period. The questionnaires were sent to recently referred patients every other month.

Data sources for attrition analysis

A representative response was required amongst the population based sample, in order that meaningful as well as more general conclusions could be made. To this purpose the Danish Statistical Institute formulated a data-set making it possible to compare responders and non-responders, in a variety of variables, gathered from varying population databases. The 1300 personal identification codes taken from the population based sample were linked to biological parents and actual family constellation. These households were defined by the following variables; parents' place of birth, type of family (one or two parents), socio-economic status of parents (including level of education and actual job profile), number of siblings, geographic position and type of dwelling. In addition, information regarding the place of birth, weight at birth of proband, and the parents' age at child-birth was available. Data was formulated by way of an anonymized database at the Danish Statistical Institute.

Socio-economic classification (SES)

Prior to analysis, SES was re-coded in accordance with guidelines formulated by the Danish Social Research Institute (66). A five step scale was used (Table 2). The two upper socio-economic classes were fused and families were classified by the status of the highest ranking parent in the household.

Clinical diagnoses in referred patients

During the Spring of 1998, clinical ICD-10 diagnoses became available from the outpatient clinic at the Child Psychiatric Department at

Table 2. Socio-economic classification (SES)

SES-class	Brief definition
Class I	academics, self employed in enterprises with more than 20 subordinate staff, and salaried employees with more than 50 subordinate staff
Class II	self employed in enterprises with 6–20 subordinate staff, salaried employees with 11–50 subordinate staff, and salaried employees with theoretical non-university level education of long duration
Class III	self employed in enterprises with 0–5 subordinate staff, owners of farms, and salaried employees with 1–10 subordinate staff
Class IV	smallholders, salaried employees having no subordinate staff and performing work not requiring expertise, and skilled manual workers
Class V	unskilled and semi-skilled manual workers, pensioners and unemployed without education

The classification is generated by the Danish Social-Research Institute (66).

Odense University Hospital. Assessment was, dependent upon age and referral diagnosis, performed by three trained specialists in Child and Adolescent Psychiatry. Diagnostic conclusions were based on a minimum of two hours of clinical interview with parents and child, evaluation of prior, or recent psychological tests, and furthermore, daycare or school observations. The clinicians were blinded to the CBCL.

Reliability data

Questionnaires were obtained from sub-samples of the clinical group, for the analysis of inter-parent and test-retest reliability.

METHODS

The instruments; CBCL, TRF and YSR

The CBCL, TRF and YSR are all similarly constructed. The checklists cover important aspects of child and adolescent psychopathology, they are known from a broad literature reference and fulfil several aims for screening and clinical assessment.

The competence items

The first part of each instrument consists of a number of competence items covering activities, social and school performance. Parents and youths are asked to specify the number of sports, hobbies, participation in organizations, jobs and friendships the proband actually takes part in. They are, also, questioned regarding the quality and degree of involvement in each specific activity. Furthermore, they are asked to describe the relationship to siblings, other children and parents, and how well they play and work by themselves. Questions concerning academic performance include current school performance, special education attendance, grade repetition and other difficulties at school. Teachers are questioned in depth as to school-performance and classroom social-functioning of the pupil. The competence aspects of the CBCL, TRF and YSR are normally scored in accordance with standardized procedure as described by Achenbach in manuals and profiles (9, 14, 15). Several authors have questioned the validity of competence scores (64) and most of the literature attaches more emphasis on problem scores. In the case of the activity scale scores, severe cultural deviance appears and the ability to differentiate between clinical and normative sub-samples is poor.

The problem items

The problem checklist comprises 118 closed and two open items, covering a wide range of behavioral and emotional problems, possibly causing concern to parents, teachers and clinicians. All three informant versions include seven closed items and one open item relating to physical, or somatic symptoms. The youths themselves are checked in much the same way, although with some exclusions, due to the irrelevance of questioning them as to more childish behavior. A 0–1–2 scale is used to score responses, these scores describe the child's behavior over the past 6 months. A 0 is scored if the description is '*not true*', a 1 if it is '*somewhat or sometimes true*', and a 2 if it is '*very or often true*'. The sum of all problem item scores is referred to as the total behavior problem score.

Crossinformant syndromes

In the validations of the CBCL and related material, including factor analysis, many authors have confirmed a number of Cross-Informant Syndromes (CIS), initially generated by Achenbach. The CIS scores are calculated by adding the scores in specific sub-groups of items. Only 85 of the 118 problem items are used to measure the eight CIS. There are three internalizing CIS named '*Withdrawn*' (9 items), '*Somatic Complaints*' (9 items) and '*Anxious/Depressed*' (14 items), which summarize to an '*internalizing score*'; three neutral CIS named '*Social Problems*' (8 items), '*Thought Problems*' (7 items) and '*Attention Problems*' (11 items); and finally, two externalizing CIS named '*Delinquent Behavior*' (13 items) and '*Aggressive Behavior*' (20 items), summarizing to an '*externalizing score*' (9).

Translation

Translation of the questionnaires was performed by professional linguistics. Initially they were translated from the original American versions into Danish, and then re-translated by a further linguist, into American. The original versions were compared with the re-translated American versions. Literally identical items were unchanged in the Danish version. Those items which were changed during the procedure were discussed within the research group and the most clinically correct and meaningful expressions were chosen for the Danish translation.

Pilot project

The CBCL, TRF and YSR, supplied with information to the participants, were tested by a smaller group of staff from the Child and Adolescent Psychiatric Department in Odense. The aim of the pilot study was to ensure that the information gathered was sufficient, to obtain an impression as to how much time parents took to complete the CBCL, to decide whether or not the TRF should be used, and finally, assist in selecting the most appropiate design for data collection. Questionnaires were distributed to 22 staff members and 17 were returned fully completed. Three had children under school age. Ten participants delivered the TRF to the classroom teacher without comment, two had negative feelings regarding the TRF (although they delivered it anyway) and a further two did not state whether they had delivered the TRF to the teacher, or not. Twelve TRF's were returned directly from teachers, of which only one was sceptical towards the project and design. The pilot phase confirmed the chosen design. Teachers were generally negative towards an alternative design, which involved returning the TRF to the family. Under such conditions, they would not be in a position to give completely honest answers. The pilot group was generally satisfied with the level of information.

Study design

A population based sample and a clinically based sample were to be compared, according to standard procedures in an epidemiological case-base design. The base is a representative sample of the study-population, from which the cases develop (i.e. the population based sample provides the opportunity to estimate item prevalence proportions, or score prevalence proportions). Cases were defined as incident referrals to child psychiatric services within the studybase in the study period ('golden standard'). The clinically based sample provides comparative prevalence proportions in referred patients. There may be a number of unidentified, potential cases in the base, which makes analyses of specificity misleading.

Analytical strategy and statistical methods

Although comparisons of Danish and other national normative results were demanded and, at the same time, development of highly predictive screening outputs were desired, analyses were made according to different strategies. Some analyses, therefore, became redundant.

Analysis of representation

Analysis of representation of the population based sample was made by chi-square tests on 2-by-2 tables, or 2-by-k tables. Responders with a Danish background were divided into a high-scoring group and a low-scoring group. A hypothesis of equal distribution of socio-demographic variables was tested in the case of both groups. Secondly, tests of equal distribution of responders and non-responders were performed on various selected variables.

Total behavior problem scores

Total behavior problem scores, internalizing and externalizing scores were handled as normally distributed. Mean total behavior problem scores were compared from the clinical and the population based samples using standard *t*-test (after testing equality of variances). Boxplots, indicating median, upper and lower quartile, and minimum and maximum scores will be presented to illustrate discrimination. Comparisons across gender, age and socio-economic status were made. Sensitivity, or correct classification of cases, due to various summarized problem scores were externally validated and, as part of the discussion, comparisons between Danish and other national mean scores will be presented.

Single item analysis

ANOVA and ANCOVA designs made by most authors within the CBCL literature will not be replicated here due to the inappropriacy of applying a model requiring normal distribution at single-item scores. Competence items, except the so-called '*activity scale*' items on the first page in the CBCL and YSR, together with all problem items, were analyzed in a logistic regression model. Discriminatory power was analyzed separately for each item. Both the competence items and problem items were dichotomized prior to analysis. All problem items were registered present when scored 1 or 2, and not present when scored 0. The varying output levels of the competence variables were dichotomized in accordance with the clinical sense. In the majority of variables a natural cutoff occurred, including average in the normal range. Cutoff levels are marked by a bold line in appendix A.1, A.2 and A.3.

Regression analysis

Odds ratios reflect the factor by which odds of being an identified and referred case *vs.* an average child from the population, are multiplied, when an item is scored present (deviant) rather than not present (normal range). The basic model is a 2-by-2 table. Characteristic 2-by-2 table outputs are presented in Fig. 1.

	Item / Symptom Present	Not present
Case	p_1 (TP)	1-p_1 (FN)
Base	p2 (FP)	1-p2 ((TN)

TP is 'true positive' test results
FP is 'false positive' test results
FN is 'false negative' test results
TN is 'true negative' test results

Sensitivity is TP/(TP+FN), i.e. the ability of the test, classifying the cases correctly
Specificity is TN/(TN+FP), i.e. the ability of the test, classifying the non-cases correctly

Odds = $P/(1-P)$, i.e. the probability that a symptom is present divided by the probability that the same symptom is not present

Odds ratio (OR) = $P_1/(1-P_1)$ / $P_2/(1-P_2)$ or $= P_1/P_2$ / $(1-P_1)/(1-P_2)$,
i.e. the odds that a symptom is present among cases divided by odds that the same symptom is present in the population
Confidence intervals for odds ratios are measured at P=0.01; OR(exp(1+/-2.57xS.E.))

The regression model equation:

$$\ln P/(1-P) = \ln R_0 + x_1 \ln R_1 + x_2 \ln R_2 + \ldots + x_k \ln R_k$$

or $P/(1-P) = R_0 \cdot R_{x1} \cdot R_{x2} \cdot \ldots \cdot R_{xk}$ (odds)

R_0: constant in the model $x_1, x_2, \ldots, x_k$ are the items from the model reported present

The probability of being a case given the item combination (test positive or T+) is:

P (case | T+) =Odds (case | T+) / 1+Odds (case | T+) (*Bayes' equation*)

A likelihood ratio can be calculated:

LR (likelihood ratio) =sensitivity/ (1-specificity)

Fig. 1. 2-by-2 table characteristics.

The single-item model included three co-variables; age with three age groups (4–5 years, 6–10 years and 11–16 years), sex (male/female) and family SES with four categories (higher-class together with higher middle-class, middle-class, lower middle-class and low-class). All odds ratios listed in appendices A.1, A.2, A.3 and B were, thereby, adjusted for the influence of these factors. Following exposure of the best discriminating items from the total data-set and from four strata (two ages strata (4–10 and 11–16 years) and the two genders) in the single-item model, a multivariate logistic regression model was applied. By forward selection and backward elimination of items (i.e. until nothing was gained in maximum likelihood), one general and four age/sex specific item-constructs emerged. Because non-correlated items were intended, the constructs were, secondarily, revised on the basis of analyses of interaction and co-linearity.

Factor analysis

Clinicians and researchers are interested in specific pathological patterns. A further approach in validity analysis included testing whether specific variables tend to cluster, or whether information can be organized in a more purposeful manner. This is one aspect of the test for construct validity. Once again, there were two available pathways of analysis, a traditional factor analysis, based on continuous variables (which has been made by other authors), or a latent trait analysis based on categorical variables. Both procedures will be demonstrated.

Traditional factor analysis

Exploratory, or principal factor analysis, as performed by Achenbach, Verhulst and Fombonne (9, 24, 25, 67), will be replicated. Data from the clinical sample are analyzed in two fractions, one including the 118 problem items from the CBCL and another including only the 89 Cross Informant Items (CII) common to the CBCL, TRF and YSR. When making factor analysis, a number of terms are essential. The 'eigenvalue' represents the amount of variance attributed to a given factor, or component. 'Factor loading' is the term for a coefficient which measures the correlation between an item and a factor. 'Rotation' is a mathematical procedure, which assists in interpretation of the factors (i.e. overlap between underlying factors can be minimized). 'Varimax rotation' is an orthogonal rotation, resulting in uncorrelated factors or components. Factors will be retained when their 'eigenvalues' are greater than one and items will, generally, be attached to a factor when its 'factor-load' is greater than 0.3. Principal component (non rotated) and varimax rotation are used.

Latent trait (i.e. latent structure) analysis

The relationship of each item to the latent continuum is modelled by a logistic, or a normal ogive curve, which is characterized by a threshold value and a slope for each item (Fig. 2) (68). The threshold for an item is the point on the latent continuum, at which there is a 50% probability of the item eliciting an 'ill' response. People above the threshold are more likely to display the symptom (i.e. the latent trait). Thresholds are equivalent to frequency of items in the population, if it is here the scale is to be used. The slope of the curve expresses the specificity of the item for the latent trait. Some symptoms are diagnostically specific, whilst others may occur in several unrelated illnesses, or even amongst normals. Slopes are equivalent to factor loadings. It is essential that the slopes or factor loadings of an item are similar in the analyses of clinical samples and community samples, in order to confirm transparency of a symptom between contexts. The latent trait model may be a one-parameter model (Rasch's models (69)) or a two-parameter model (Birnbaum et al's model (70)). In the Rasch model, the slopes of the logistic curves are assumed equal for all items, and are characterized only by the threshold parameter. In the Birnbaum model the slope, as well as the threshold, may be different for each item. In this piece of research, however, only the presumptions of the model will be discussed without performing the actual test. Factor load and frequency are the only terms used.

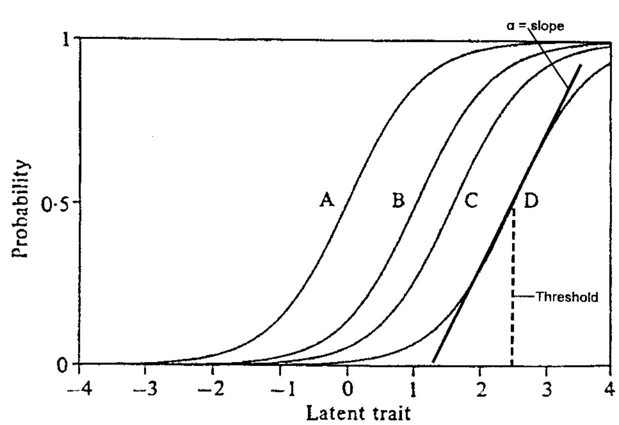


Fig. 2. The latent trait model. Each curve represents an item, characterized by a threshold, i.e. the point on the latent continuum at which there is a 50% probability of the item eliciting an 'ill' response, and a slope, i.e. expressing the specificity of the item for the latent trait.

The 118 dichotomized problem item scores from referred and non-referred children are fitted into models with varying numbers of dimensions, by which the variance of the information must be explained. Models including from one to eight dimensions will be undertaken in this exploratory analysis. Items with factor loadings less than 0.6 (on a single dimension 0.5) on a dimension are excluded. Single items with borderline factor loadings can be included due to clinical relevance, and a number of latent structures, or scales, can be formed. Each latent structure model is secondly confirmatory analyzed. Factor loadings are compared between clinical and community settings and items are ranked in accordance with frequency in the population based sample. It is essential that each scale is clinically relevant (content validity) and that each includes variables with symptom progression, to ensure that scale-score reflects severity and scale-score cutoff can be effective. The results of the various factor analyses are compared and components, or latent trait scales, are externally validated as predictors of clinical diagnoses in the referred sample.

Reliability

Test-retest, inter-parent reliability and comparisons of different raters (parents, teachers and youths), are normally tested with correlation statistics. Pearson correlation coefficient, or the like, provides us with little information and is, therefore, calculated only for comparative use. Instead, reliability will be examined from sum-difference plots.

Software used

SPSS for Windows, SAS for Windows (principal component analysis), NOHARM (latent trait analysis).

Ethics

Along with the questionnaires, families and teachers received information relating to the standardization project. Participants were guaranteed anonymity. By delivering the TRF to the school, the parents and child approved that the classroom-teacher returned the checklists directly to the project group. Two mailed reminders were allowed by the Danish ethical committee, by whom the entire study design was ratified. Furthermore, the project was sanctioned by ministerial register authorities.

RESULTS

Participation rates

The population based sample

In the population based sample, three out of the 1300 children were unknown at the address. No substitutes were included. Within two weeks of distribution, 647 (or 50%) had replied. The first reminder resulted in a further 100 replies, and the second reminder increased the number of responders to 798 (61.4%). A total number of 779 (59.9%) replied to the CBCL, 547 (49.7%) responded the TRF and 355 (55.8%) replied to the YSR (71). Participation declined with age-increase and was lowest amongst boys. Highest participation was found amongst four to five years old girls (73%).

Children of non-Danish background, defined as families where both parents were born outside of Denmark, represented 77 cases (5.9%). Thirty-five (45.5%) responses were received from this group, of which some were insufficient. The informers at the CBCL were 647 (83%) mothers (or mother and father combinations). In 111 cases (14%) the father replied and in 12 cases (2%) responders were alternative persons with a relationship to the child. On nine occasions, the informer was unknown.

The clinically based sample

In the clinical sample, the questionnaires were mailed to 302 probands, 203 boys and 99 girls, referred to the child psychiatric services. Material was received from 157 (52%) of the families involved. The CBCL were answered in 146 cases. TRF was returned in 118 cases and YSR in 58 cases. Fewer fathers (5%) and more alternative-carers (6%) tended to be the informers on the CBCL in the clinically based sample.

Analyses of representation

In order to illustrate the representation of the responding groups, all possible data sources were used. When indexing family SES, two-parent families were SES-classified, according to the highest ranking parent, in accordance with the previously mentioned definitions. In Table 5, the family SES distribution from population registers are listed, together with family SES scored from personal information provided by the responders.

Table 3. Response rates in the population based sample (N = 1300)

Age	Boys		Girls		Total	
years	N	%	N	%	N	%
4–5	71	71.0	73	73.0	144	72.0
6–10	148	59.2	155	62.0	303	60.6
11–14	108	54.0	144	72.0	252	63.0
15–16	38	38.0	61	61.0	99	49.5
Total	365	56.2	433	66.6	798	61.4

Table 4. Response rates in the clinically based sample (N = 302)

Age	Boys		Girls		Total	
years	N	%	N	%	N	%
4–5	11	40.7	1	14.3	12	35.3
6–10	60	56.1	16	48,5	76	54.3
11–14	33	63.5	17	53.1	50	59.5
15–16	8	47.1	11	40.7	19	43.2
Total	112	55.2	45	45.5	157	52.0

The population based sample

Responders from the population based sample and the clinically based sample were unequally distributed according to family SES. Although family SES distribution of responders in the clinical sample appeared to be practically equal to population figures, it is possible that the sample deviated from an unselected clinical group. Family SES group 1+2 and 5 were, respectively, over and under represented amongst responders in the population based sample, compared with background distribution.

For the drop-out analyses made at the Danish Statistical Institute, responders in our population based sample were divided into two groups (71). A total of 59 participating children (7.8%), with at least one Danish parent, had total behavior problem scores above the 95 percentile, on at least one questionnaire, these were placed in the high-scoring group. The remaining 702 children constituted a low-scoring group. Children from one-parent families were most frequently placed in the high-scoring group (chi square = 15.9; (df = 1); $P<0.01$). Family SES were not, significantly, unequally distributed (chi square = 6.0; (df = 4); $P=0.20$), although there was a tendency toward a

Table 5. Socio-economical status (SES) in families. Distribution in the population and among responders

Family	Population		Responders		Responders	
SES*	DK[a]	Fynen[a]	Norm (N = 772)		Clinic (N = 140)	
group	%	%	N	%	N	%
1+2	16.0	17.3	196	25.2	21	14.8
3	31.0	29.8	192	24.6	38	26.8
4	30.9	30.3	265	34.0	42	29.6
5	22.1	23.1	119	15.3	39	27.5

*Highest ranking parent (stepparent) according to outlines from the Danish Social-Research Institute (Table 2); [a]demographic report from Fynen (65).

higher total behavior problem score in lower SES groups. Type of family dwelling (rented- house/ apartment *vs.* owned) and geographical situation (urban *vs.* rural), parents' age at child-birth, number of siblings, as well as the child's weight at birth showed no significant influence on total behavior problem score on the CBCL. In the secondary phase, a hypothesis of equal distribution of responders and non-responders was tested. An even distribution of single-parent families and two-parent families was found in female proband families. Although the boys of single parents were, significantly, less represented in the responding group (chi square $=3.7$; (df$=1$); $P=0.05$).

The clinically based sample

The 157 responders in the clinically based sample were compared with the total of 302 referred children included in the clinical study group. The distribution of age and gender was equal. Forty-seven of the referred children did not show up for clinical assessment. They were unequally distributed, with 33 in the non-responding group *vs.* 14 in the responding group. Assessed children were divided into major diagnostic groups in accordance with the ICD-10 manual (5). The following list of diagnostic categories were used; organic mental-disorders (F00-F09), mental and behavioral disorders, due to psychoactive substance abuse (F10-F19), schizophrenia, schizotypal and delusional disorders (F20-F29), affective disorders (F30-F39), phobic-anxiety disorders (F40), obsessive-compulsive disorders (F42), adjustment disorders (F43), dissociative disorders (F44), other neurotic disorders (F48), eating disorders (F50.0–9), specific personality disorders (F60), mental-retardation (F70-F79), speech disorders (F80.0–9), disorders of academic skills (F81.0–9), disorders of motor function (F82.0–9), mixed developmental disorders (F83.0–9), pervasive developmental disorders (F84.0–9), unspecified disorders of psychological development (F89.0–9), hyperkinetic disorders (F90.0–9), conduct disorders (F91.0–9), mixed conduct and emotional disorders (F92.0–9), emotional disorders (F93.0–9), disorders of social functioning (F94.0–9), tic disorders (F95.0–9) and other emotional and behavioral disorders (F98-F99). In the assessed group of 255 referred children, seventeen, all in the responding group, did not fulfil criteria for any ICD-10 diagnosis. The remaining 238 children, 161 boys and 77 girls had main clinical diagnoses with a distribution demonstrated in Table 6.

The number of probands in most diagnostic groups was limited. Differences in distribution of responders and non-responders within diagnostic groups were not statistically significant, although it would appear that children and adolescents with

Table 6. ICD-10 diagnoses in the assessed clinical sample (N=238). Only the primary psychiatric diagnosis is included

		Responders		Non-responders	
ICD-10 code	ICD-10 diagnosis (short text)	boys N (%)	girls N (%)	boys N (%)	girls N (%)
F00-F09	Organic mental disorder	1 (1.1)	1 (2.6)	—	—
F10-F19	Mental and behavioral disorder due to psychoactive substance use	—	—	—	2 (5.3)
F20-F29	Schizophrenia, schizotypal and delusional disorder	—	1 (2.6)	3 (4.1)	1 (2.6)
F30-F39	Affective disorder	1 (1.1)	—	—	1 (2.6)
F40	Phobic anxiety disorder	—	1 (2.6)	—	—
F42	Obsessive-compulsive disorder	4 (4.6)	1 (2.6)	1 (1.4)	4 (10.5)
F43	Reaction to severe stress, adjustment disorder	5 (5.7)	5 (12.8)	7 (9.5)	2 (5.3)
F44	Dissociative disorder	—	—	—	1 (2.6)
F48	Other neurotic disorder	1 (1.1)	3 (7.7)	2 (2.7)	1 (2.6)
F50.0–9	Eating disorder	1 (1.1)	7 (17.9)	—	7 (18.4)
F60	Specific personality disorder	—	1 (2.6)	2 (2.7)	—
F70-F79	Mental retardation	1 (1.1)	—	4 (5.4)	1 (2.6)
F80.0–9	Disorder of speech and language	1 (1.1)	1 (2.6)	2 (2.7)	—
F81.0–9	Disorder of scholastic skills	1 (1.1)	—	2 (2.7)	1 (2.6)
F83.0–9	Mixed developmental disorder	4 (4.6)	—	2 (2.7)	1 (2.6)
F84.0–9	Pervasive developmental disorder	20 (23)	3 (7.7)	14 (19)	1 (2.6)
F89.0–9	Unspecified disorder of psychological development	—	—	1 (1.4)	1 (2.6)
F90.0–9	Hyperkinetic disorder	4 (4.6)	—	4 (5.4)	1 (2.6)
F91.0–9	Conduct disorder	26 (30)	4 (10.3)	15 (20)	—
F92.0–9	Mixed disorder of conduct and emotions	3 (3.4)	1 (2.6)	—	—
F93.0–9	Emotional disorder	7 (8)	6 (15.4)	5 (6.8)	7 (18.4)
F94.0–9	Disorder of social functioning	2 (2.3)	4 (10.3)	7 (9.5)	6 (15.8)
F95.0–9	Tic disorder	—	—	2 (2.7)	—
F98-F99	Other emotional and behavioral disorder	5 (5.7)	-	1 (1.4)	—
Total		87 (100)	39 (100)	74 (100)	38 (100)

severe mental disorders (e.g. schizophrenia, mental disorders due to psychoactive substance-abuse, mental retardation and disorders of social functioning) tended to be less-represented amongst responders. This fact negatively influences the value of the factor analysis presented later, owing to fewer delusional and thought disturbing symptoms in the item pool. Although low prevalent (according to Table 6), developmental speech-disorders, disorders of academic skills and disorders of motor function were frequently found to be comorbid. Item responses reflecting most child psychiatric pathology should, fundamentally, be represented in the responding group.

Total competence score

In the original manual, a total competence score is calculated in accordance with a specific algorithm and expressed in a so-called T-score. This T-score was not replicated in the currrent study because of an opaque strategy used. Competence items were evaluated individually.

Table 7. Mean values and standard deviations for total behavioral problem scores

			Population sample				Clinical sample	
Sex	Source	Age in years	n	Mean value	s.d.	Collapsed mean value (s.d.)	n	Collapsed mean value (s.d.)
Boys	CBCL	4–5	71	18.9	13.4	19.7	102	60.8
		6–10	146	20.9	17.3	(16.5)		(27.8)
		11–16	141	19.0	17.0			
	TRF	6–10	123	20.7	23.4	20.4	87	62.4
		11–16	119	20.0	20.5	(22.0)		(33.0)
	YSR	11–16	140	30.0	19.6	30.0 (19.6)	32	47.3 (25.4)
Girls	CBCL	4–5	73	16.1	10.5	16.0	44	50.5
		6–10	153	15.6	11.8	(13.2)		(34.6)
		11–16	195	16.4	15.1			
	TRF	6–10	131	12.1	14.5	14.4	30	39.3
		11–16	174	16.1	19.3	(17.5)		(26.0)
	YSR	11–16	195	31.7	18.6	31.7 (18.6)	25	51.9 (34.8)

CBCL 4–16 mean total score population, boys and girls collapsed was 17.7 (s.d. 14.9), N = 779.

TRF 6–16 mean total score population, boys and girls collapsed was 17.0 (s.d. 19.8), N = 547.

YSR 11–16 mean total score population, boys and girls collapsed was 31.0 (s.d. 19.1), N = 335.

Total behavior problem score

Total behavior problem score is one of the main output parameters in the majority of literature regarding the CBCL and related material. The total behavioral problem score of the CBCL is calculated by adding all problem item scores with two exceptions; item 2: *asthma* and item 4: *allergy*. Scores range from zero to 232. In the TRF, total behavior problem score is calculated by adding all 118 TRF problem item scores (range; 0–236). The YSR total behavior problem score is obtained by adding 100 of the item scores, omitting 16 positive items (numbers 6, 15, 28, 49, 59, 60, 73, 78, 80, 88, 92, 98, 106, 107, 108 and 109), which have substitute items unsuitable for self-reporting, together with item 2: *asthma* and item 4: *allergy* (range; 0–200) (15).

Total behavior problem scores in the population based sample (stratified in boys and girls and in the age groups 4–5 years, 6–10 years and 11–16 years), showed practically identical patterns of distribution illustrated in a histogam in Fig. 3. Mean values and variances were calculated and presented in Table 7. Means and standard deviations (s.d.) of the total behavior problem scores in the clinical sample are also presented in Table 7.

Case vs. base sample discrimination

Mean value of total behavior problem score in the CBCL were compared from the population, and from the clinically based samples. T-test for equality of means (unequal variances) was used. The two mean values were 17.7 and 57.7, respectively, and the test provided a t-value of 15.57 (df = 158.4) indicating a significantly high difference ($P<0.0001$). A comparison of means in total behavior problem scores in the two samples, derived

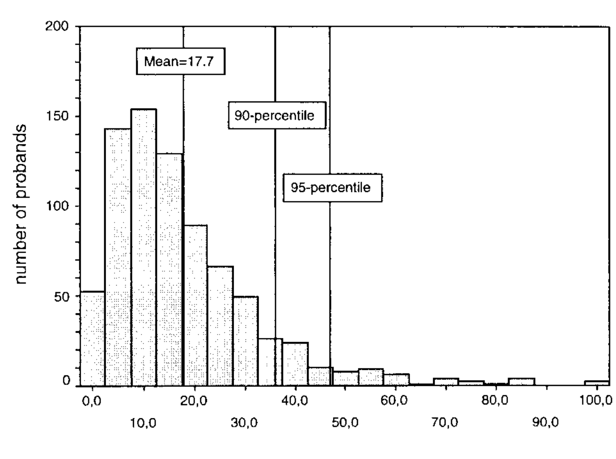


Fig. 3. Distribution of total behavior problem score at the CBCL amongst responders in the population based sample (N = 779).

Supplementum

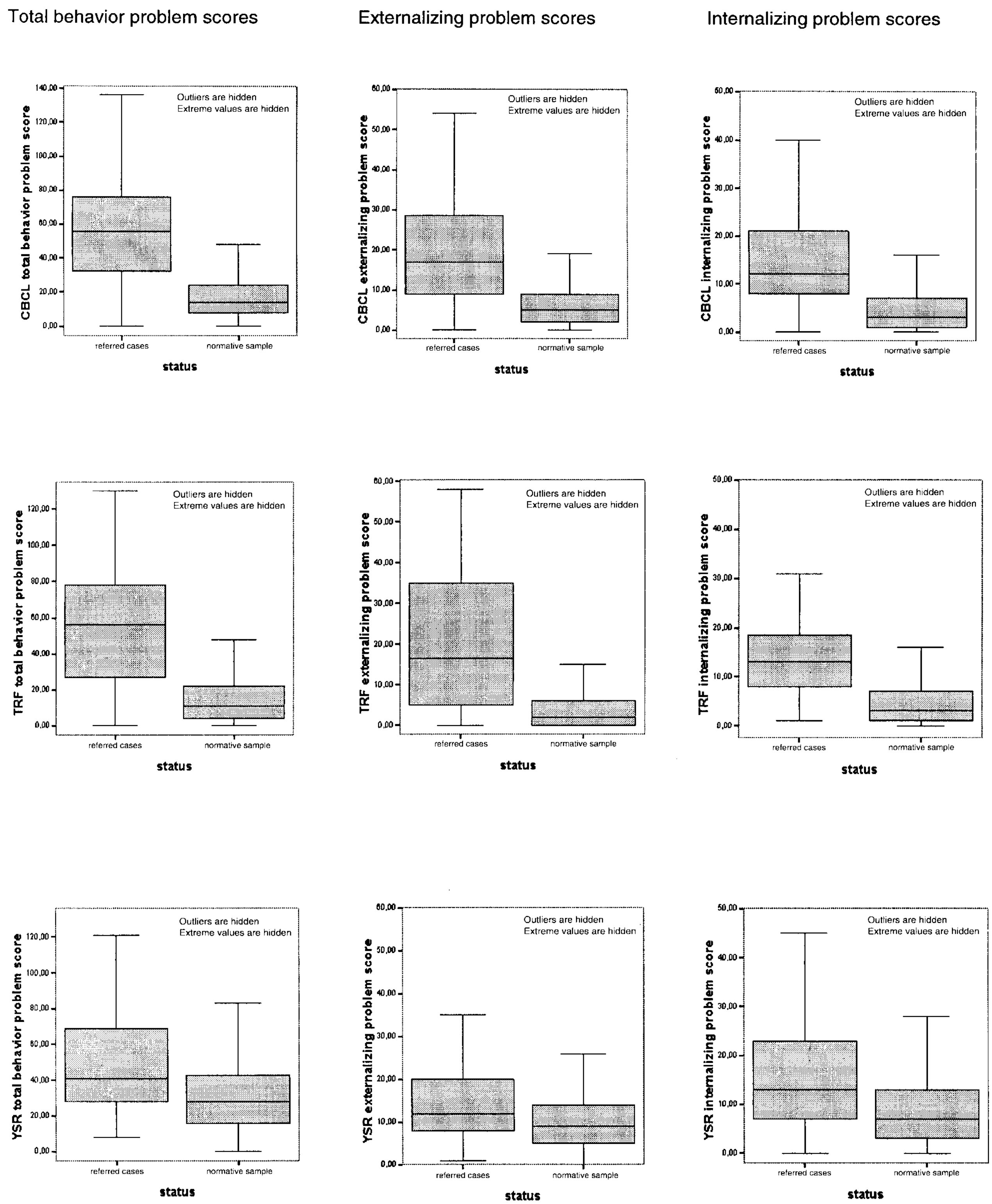


Fig. 4. Boxplots. Distinction between behavioral scores. Normative *vs.* referred samples.

from TRF and YSR respectively, provided results similar to that of the CBCL. T-values at 12.52 (df = 134.6) and 4.50 (df = 64.1) both indicated significantly higher mean scores in the case of referred children.

In order to illustrate the ability of the CBCL-, TRF- and YSR total behavior problem scores to distinct cases from baseline population, a number of boxplots are demonstrated in Fig. 4. A boxplot indicate the median score, the upper and lower quartile (25- and 75 percentile), and the maximum and minimum scores.

Sex differences

Boys in the population based sample scored higher than girls. Mean total behavior problem scores in the CBCL was 19.7 for boys and 16.0 in the case of girls. This provided a t-value of 3.39 (df 681.2; $P<0.001$). Mean values in TRF were 20.4 for boys and 14.4 for girls; $t=3.46$ (df 452.4; $P<0.001$). Furthermore, both parents and teachers scored boys higher than girls in the clinical sample. In the CBCL, mean total problem scores were 60.8 and 50.5 respectively, at TRF scores were 62.4 and 39.3. Only the difference at TRF was statistically significant. In self-reported behavior problems (YSR) no significant sex differences whatsoever were present.

Age differences

When comparing mean problem scores for younger and older children, 4–10 years *vs.* 11–16 years, no differences were found in the population based sample. In the referred sample, differences occurred at $P<0.05$ level, with younger children scoring higher than older children.

SES differences

For this analysis family SES groups were recoded into two SES levels; 'high' including SES groups I, II and III and 'low' including SES groups IV and V, according to Table 2. A t-test was performed, comparing mean total problem score in the population based sample for the two SES levels. High SES level children had CBCL mean total problem score on 14.8 (s.d. 12.2) *vs.* 20.4 (s.d. 16.4) in low SES level children. The t-value was 5.34 ($P<0.0001$). Furthermore, at the TRF a significantly high difference was found between the two SES levels, although no significant difference was found in self reported mean total problem score.

Table 8. Internalizing scores: Mean values and standard deviations

			Population			Clinical
Sex	Source	Age in years	Mean value	s.d.	Collapsed mean value (s.d.)	Collapsed mean value (s.d.)
Boys	CBCL	4–5	3.4	3.8	4.8	15.4
		6–10	5.1	5.8	(5.4)	(10.0)
		11–16	5.4	5.5		
	TRF	6–10	5.5	6.8	4.8	13.9
		11–16	4.2	5.2	(6.1)	(10.0)
	YSR	11–16	7.7	6.6	7.7 (6.6)	11.7 (7.9)
Girls	CBCL	4–5	3.2	3.2	4.4	16.6
		6–10	4.2	3.8	(4.2)	(12.2)
		11–16	5.0	4.8		
	TRF	6–10	5.3	6.1	4.9	(16.9)
		11–16	4.6	5.2	(5.6)	(8.1)
	YSR	11–16	9.8	7.2	9.8 (7.2)	20.7 (14.0)

Internalizing and externalizing behavior scores

Internalizing and externalizing behavior scores were calculated in accordance with the US manuals. Normative and clinical scores are listed in Tables 8 and 9. Discrimination is illustrated by boxplots in Fig. 4. In the population based sample, and even more so in the clinically based sample, boys were more externalizing than girls. No significant age effect was found in externalizing behavioral scores.

Table 9. Externalizing scores: Mean values and standard deviations

			Population			Clinical
Sex	Source	Age in years	Mean value	s.d.	Collapsed mean value (s.d.)	Collapsed mean value (s.d.)
Boys	CBCL	4–5	6.2	4.7	7.4	21.4
		6–10	8.2	6.6	(6.4)	(12.5)
		11–16	7.0	6.7		
	TRF	6–10	6.5	9.4	6.5	24.3
		11–16	6.5	8.1	(8.8)	(16.9)
	YSR	11–16	11.0	6.8	11.0 (6.8)	16.3 (10.4)
Girls	CBCL	4–5	6.1	5.1	5.4	14.4
		6–10	5.5	4.9	(5.2)	(12.3)
		11–16	5.3	5.4		
	TRF	6–10	2.8	4.9	3.9	9.6
		11–16	4.7	7.4	(6.5)	(13.7)
	YSR	11–16	9.7	5.9	9.7 (5.9)	13.2 (8.9)

In contrast, however, internalizing behavior was almost equally reported by parents and teachers, in the population based sample. Internalizing behavior produced a tendency towards age-increase in the ratings of parents. Amongst girls in particular a high discrepancy between self-rating and parents/teachers rating of internalizing problems was found.

Competence item analysis

Competence items at the CBCL and YSR consist of three main categories; 1) activity items, which were not analyzed in this study, 2) social competence items, and 3) school performance items. The distribution of scores in the population based sample and in the clinically based sample are shown in appendices A.1 and A.2 for social competence and school competence, respectively. All six social competence items were, significantly, positive discriminators between the two samples in parents-rating at the CBCL. Only 5% of children in the population based sample had *none*, or *one single good friend* (item V.1/CBCL) compared with 40% in the clinically based sample. Discriminative power was calculated, in accordance with the logistic regression model referred to in the chapter 'statistical methods'. The adjusted odds ratio with confidence intervals for item V.1/CBCL was 15.40 (7.86–29.91; $P=0.01$). The items VI.b/CBCL: *how well does the child get along with other kids* and VI.c/CBCL: *how does the child behave with his/her parents* produced odds ratios at 38.72 (13.21–114.37; $P=0.01$) and 12.11 (3.70–39.34; $P=0.01$), respectively. The three questions comprised the best discriminating CBCL social competence items (appendix A.1).

When the youths rated the same items at the YSR themselves, only one item was found to be discriminatively significant. Item VI.c/YSR: *how well do you get along with your parents*, discriminated with an odds ratio of 7.69 (1.52–38.83; $P=0.01$). Item V.2/CBCL: *number of contacts with friends per week outside school hours*, were scored significantly higher in the case of older children, and girls were significantly better at *getting along with brothers and sisters* (item VI.a/CBCL) than boys.

School competence rating was a strong discriminator in the hands of teachers. In all of the seven items in the evaluation of academic performance and classroom attitude (appendix A.3), the clinical probands scored, significantly, more frequently below average compared with the population based probands. Best discriminating competence item was VIII.2/TRF: *compared to typical pupils of the same age, how appropriately is he/she behaving*, with an OR at 9.03 (4.51–18.06; $P=0.01$). In the three academic performances included, teachers scored children from high SES families as being more competent. Girls were rated significantly more often, at or above average grade, in reading and spelling, than boys. Girls were, also, more industrious, behaved better and had higher learning ability than the boys in the classroom. In the case of parents and youths rating, academic performance was far less discriminatory (appendix A.2). Parents, also, rated their daughters higher than their sons in reading and in spelling. Reading, spelling and arithmetic performances (VII.a-c/CBCL) all demonstrated odds ratios with confidence intervals slightly above one, indicating little, although significant use in the screening of deviant children from a random sample.

Behavioral and emotional problem item analysis

Case vs. base sample discrimination

A total of 118 behavioral and emotional problem items are found in the CBCL, TRF and YSR. Appendix B displays detailed frequency data from the collapsed sample and eight subgroups; boys 4–10 years, boys 11–16 years, girls 4–10 years and girls 11–16 years, each divided into population based and clinically based data. From looking at percentage distribution on the three step (0–1-2) item scale, it was apparent that summarized item scores were not normally distributed. The item scores derived from the clinical sample, in particular, showed asymmetrical distribution. Discrimination analyses were made exclusively on dichotomized data, and the results of the single item logistic regression analyses are listed in appendix B relating to all three questionnaires.

Poor discriminators

Prior to examination of the subgroups, the regression model was applied to behavioral problem item data taken from the total number of responders. Single item discriminative power between population and referrals was demonstrated by an odds ratio (recognized as 'collapsed OR' in appendix B), adjusted for the influence of confounders (SES, age and sex). Two items, 78/CBCL: *smears feces* and 110/CBCL: *wishes to be opposite sex*, were reported present so rarely in both the clinical and population based samples, that statistical power was insufficient to show any significant difference. Seven items, numbered 15, 18, 67, 72, 73, 82 and 106 at the CBCL, had prevalence rates in the population of less than 2%. These symptoms were more frequently present in the clinical population to such a degree that statistical power was sufficient, although

Table 10. CBCL emotional and behavioral problem items. Items that do not discriminate referred children from general population

Item (short text)		Item (short text)	
2	allergy	59	plays with private parts in public
3	argues a lot	63	prefers older kids
4	asthma	71	self-conscious
5	acts like opposite sex	77	sleeps much
28	eats nonfood	78	smears BM
32	needs to be perfect	96	thinks about sex too much
44	bites fingernails	99	too concerned with neat/clean
55	overweight	105	uses alcohol or drugs
56d	eye problems	110	wishes to be opposite sex
56e	skin problems		

All items provided OR above 1.0 but lower confidence values were below 1.0 ($P = 0.01$).

escorted with low reliance. The discriminative odds ratios for these items were extremely high, although confidence intervals for OR at $P = 0.01$, were substantial. A total of 19 behavioral and emotional problem items in the CBCL failed to discriminate referred children from the background population. These are listed in Table 10.

Positive discriminators at the CBCL

The best discriminating items at the CBCL behavioral and emotional problem section were ranked and listed in Table 11. The CBCL problem items 82: *steals outside home* and 15: *cruel to animals* had extremely high odds ratios, whilst items 25: *doesn't get along*, 91: *suicidal talk*, 103: *unhappy, sad or depressed*, 84: *strange behavior*, 13: *confused* and 48: *not liked* additionally were associated with lower confidence level for OR above seven ($P = 0.01$).

Table 11. CBCL emotional and behavioral problem items. Best discriminators

Item (short text)		Odds ratio	Confidence intervals*
82	steals outside home	21.63	4.61–100.69
25	doesn't get along	19.04	10.31–35.40
15	cruel to animals	17.20	4.66–64.12
91	suicidal talk	16.34	7.15–37.05
103	unhappy, sad, depressed	16.14	8.70–29.87
84	strange behavior	15.92	7.77–32.77
13	confused	15.29	7.86–29.91
48	not liked	14.07	7.37–26.64
66	repeats actions (compulsions)	12.96	4.75–35.24
106	vandalism	12.77	4.70–34.89
85	strange ideas	12.75	5.77–28.41
18	harms self	12.71	3.99–40.31
80	stares blankly	12.03	5.72–25.41
45	nervous	11.61	6.58–20.40
67	runs away from home	11.43	3.70–35.55
30	fears school	10.44	4.38–25.12
21	destroys others things	10.39	4.68–23.03
100	trouble sleeping	10.24	5.41–19.54
46	nervous movements	10.19	4.83–21.44

*All confidence intervals are based on *P*-values at 0.01.

Table 12. TRF emotional and behavioral problem items. Best discriminators

Item (short text)		Odds ratio	Confidence intervals*
91	suicidal talk	40.56	4.44–368.78
21	destroys others things	22.13	6.30–78.20
66	repeats actions (compulsions)	20.48	3.39–123.84
18	harms self	14.10	2.40–83.37
84	strange behavior	12.00	5.52–25.82
83	stores up unneeded things	11.37	2.76–46.69
103	unhappy, sad, depressed	10.98	5.51–22.06
20	destroys own things	10.90	4.22–28.24
9	can't get mind off thoughts (obsessions)	10.42	3.91–27.57

*All confidence intervals are based on *P*-values at 0.01.

These problem items comprised particularly good discriminators.

Discriminative value of the TRF problem items

Amongst the teacher-reported behavioral problems, nine items discriminated the two samples with $OR > 10.0$ (Table 12). Six of these (number 18, 21, 66, 84, 91 and 103/TRF) were, also, represented amongst the best discriminating items in the corresponding parent rating. Twenty items of the 118 problem items at TRF did not discriminate between population or clinical group.

Discriminative value of the YSR problem items

There was a tendency amongst adolescents to report more symptoms relating to themselves, than parents and teachers. Owing to the relatively limited sample sizes, and the low response rates, the analyses were made on a total of 57 referred youths and 335 youths from the population based sample; it was difficult to illustrate significant differences due to referral status. In 24 items of the 89 cross-informant items, referred youths rated themselves, significantly ($P < .01$) higher than youths from the background population. Only seven emotional and behavioral problem items discriminated with $OR > 5.0$ (Table 13). Four items reoccurred at the CBCL best discriminator list, these were items numbered 18, 66, 67 and 82/YSR.

Table 13. YSR emotional and behavioral problem items. Best discriminators

Item (short text)		Odds ratio	Confidence intervals*
67	runs away from home	11.04	2.02–60.11
82	steals outside home	9.90	1.96–49.85
56g	vomiting	7.21	1.59–32.99
81	steals at home	6.67	1.59–28.20
66	repeats actions (compulsions)	6.24	1.82–21.40
18	harms self	5.38	1.21–23.82
30	fears school	5.21	1.48–18.34

*All confidence intervals are based on *P*-values at 0.01.

Supplementum

Table 14. CBCL emotional and behavioral problem items. Boys, 4–10 years. Best discriminators

Item (short text)		Odds ratio	Confidence intervals*
80	stares blankly	37.2	7.2–193.4
25	doesn't get along	21.3	8.5–53.8
67	runs away from home	21.1	2.8–160.8
21	destroys others things	20.3	6.0–67.2
106	vandalism	19.0	4.8–73.9
13	confused	18.7	6.2–54.6
15	cruel to animals	18.1	3.3–99.1
103	unhappy, sad or depressed	17.7	6.1–50.6
48	not liked	15.3	5.8–40.7
91	suicidal talk	14.9	4.6–48.5
100	trouble sleeping	14.3	4.9–42.1
45	nervous	13.0	5.3–32.1
8	can't concentrate	12.3	4.5–33.2
84	strange behavior	11.6	4.0–33.2
46	nervous movements	10.6	3.6–31.2
72	sets fire	10.3	2.4–44.5
82	steals outside home	10.2	1.5–68.2

*All confidence intervals are based on *P*-values at 0.01.

Sex differences

At the CBCL, parents scored boys significantly higher than girls in 30 problem items, all marked with a 'B' in appendix B. Girls were scored higher, marked 'G', on two items; 5: *acts like opposite sex* and 71: *self-conscious*. At the TRF, 41 items scored significantly higher in the case of boys, and 6 items scored higher in girls. At the YSR girls scored themselves higher ($P<0.01$) on nine items; number 17: *daydreams*, 18: *harm self*, 27: *jealous*, 56e: *skin problems*, 71: *self-conscious*, 75: *shy*, 87: *moody*, *110*: *wishes to be opposite sex* and 112: *worries*. Boys scored themselves higher on eight items.

Table 15. CBCL emotional and behavioral problem items. Boys, 11–16 years. Best discriminators

Item (short text)		Odds ratio	Confidence intervals*
30	fears school	45.5	2.8–750.9
91	suicidal thoughts	39.5	4.2–370.1
84	strange behavior	16.6	3.6–77.6
103	unhappy, sad or depressed	16.5	4.8–56.5
45	nervous	14.6	4.2–50.1
56c	nausea, feels sick	13.1	2.2–77.7
33	feels unloved	13.0	4.0–42.6
25	doesn't get along	12.3	3.8–40.1
40	hear things that aren't there	11.9	1.2–114.6
100	trouble sleeping	10.9	2.9–41.5
95	temper tantrums	10.7	3.2–35.8
19	demands attention	10.7	3.0–38.7
54	overtired	10.2	3.1–33.2
56a	aches, pains	10.2	2.1–48.8
81	steals at home	10.1	1.7–60.9

*All confidence intervals are based on *P*-values at 0.01.

Age differences

On 27 problem items at the CBCL younger children (between 4 and 10 years), scored higher. These are marked with an 'Y' underneath the total odds ratios in appendix B. Twenty problems were more frequent amongst older children (between 11 and 16 years), marked with 'O'. At the TRF, only one problem item, 11: *too dependent*, was reported significantly more frequently in younger children. Two items, 59: *sleeps in class* and 107: *dislikes school*, were reported more frequently amongst older children by classrom teachers. The YSR was analyzed in one age group only.

SES differences

Thirty-three emotional and behavioral problem items were rated significantly higher by parents in the low family SES-group (marked with an 'S' in appendix B). The same tendency was present in eight TRF problem items and in four YSR problem items. Two items, only, number 8: *can't concentrate* and 34: *feels persecuted*, showed consensus in SES influence between the CBCL and TRF. No single problem item, whatsoever, was rated significantly higher in upper SES-group children.

Alternative screening constructs from the CBCL problem item pool

A study of discriminative power within sub-groups contributed new information. Tables 14 to 17

Table 16. CBCL emotional and behavioral problem items. Girls, 4–10 years. Best discriminators

Item (short text)		Odds ratio	Confidence intervals*
25	doesn't get along	>100	$26.5 \rightarrow >1000$
48	not liked	>100	$13.0 \rightarrow >1000$
85	strange thoughts	>100	$9.1 \rightarrow >1000$
46	nervous movements	>100	$5.4 \rightarrow >1000$
62	clumsy	51.0	4.5–570.1
13	confused	46.1	7.1–300.7
84	strange behavior	40.2	7.0–229.9
23	disobeys at school	40.2	4.0–404.6
73	sex problems	38.1	$1.1 \rightarrow >1000$
16	mean to others	34.4	4.2–283.6
37	fighting	32.5	5.0–211.9
30	fears school	31.3	4.2–233.8
91	suicidal thoughts	30.9	4.3–223.4
103	unhappy, sad or depressed	29.0	5.6–150.6
111	withdrawn	28.3	3.4–232.2
35	feels worthless	25.7	5.3–122.4
80	stares blankly	24.3	3.6–162.7
19	demands attention	24.3	1.9–317.3
8	can't concentrate	23.5	3.8–146.2
45	nervous	20.3	4.2–96.3
61	poor school work	20.1	3.8–106.8

* All confidence intervals are based on *P*-values at 0.01.

represent the best discriminating items in every strata, all with adjusted odds ratios greater than 10. Owing to a large number of highly discriminating items in the group of girls (4–10 years), Table 16 includes items with OR > 20 and 99%-confidence intervals above 1.

The items in Tables 11, 14–17, respectively, were input in a multivariate logistic regression model. Prior to entering the model, items that provided odds ratios with lower confidence levels below 2 ($P=0.01$) were excluded. This parallels increasing demands for statistical power in the single item analysis to $P=0.001$, or even more. Multivariate logistic regression with forward selection ($P<0.1$) and backward elimination ($P>0.1$) was performed. In the age and gender stratified analyses, four different item combinations of three to six problem items developed.

The collapsed sample

After running the multivariate logistic regression analysis with the 19 items from Table 11, eight items fitted the model. Item number 106: *vandalism*, had weighty interactions with three other items, 13: *confused*, 25: *doesn't get along* and 103: *unhappy, sad or depressed*. The item was expelled from the solution. The remaining seven items, together with regression coefficients (B), significance levels (*P*) and odds ratios (OR or exp (B)) are listed in Table 18. Furthermore SES influence, as well as the constant in the model, is reflected in the table.

Boys, 4–10 years

Six items from Table 14 remained after the analysis. Interaction was found between item 8: *can't concentrate* and item 21: *destroys others things*. The latter was excluded and the result was a five item solution, seen in Table 18.

Table 17. CBCL emotional and behavioral problem items. Girls, 11–16 years. Best discriminators

Item (short text)		Odds ratio	Confidence intervals*
67	runs away from home	34.4	1.4–856.2
20	destroys own things	26.9	2.3–308.4
85	strange thoughts	25.4	2.6–251.5
84	strange behavior	17.3	2.3–131.7
103	unhappy, sad or depressed	17.1	4.5–65.1
13	confused	16.4	3.3–83.0
89	suspicious	14.2	3.3–61.2
18	harms self	13.9	1.8–105.7
40	hear things that aren't there	12.9	1.5–109.2
46	nervous movements	12.3	1.8–84.6
50	too fearful or anxious	11.6	3.1–43.0

* All confidence intervals are based on *P*-values at 0.01.

Boys, 11–16 years

Five of the 13 items in Table 15 fitted the model. By trying all different combinations, interactions were found between item 25: *doesn't get along*, and items 45: *nervous* and 100: *trouble sleeping*. Item 25 was excluded. Interaction was, also, found between item

Table 18. Screening constructs derived from the CBCL problem item pool. Multivariate logistic regression analyses

Item (short text)		B*	Significance	Exp (B)
Boys, 4-10 years				
8	can't concentrate	1.4184	.0036	4.1306
25	doesn't get along	2.1592	.0000	8.6642
45	nervous	1.3008	.0066	3.6723
80	stares blankly	2.1519	.0038	8.6015
91	suicidal talk	1.4375	.0221	4.2101
	social group	0.0346	.8405	1.0352
	constant	−3.7901	.0000	
Boys, 11–16 years				
45	nervous	2.3742	.0000	10.7421
91	suicidal talk	3.4433	.0002	31.2915
100	trouble sleeping	1.3694	.0337	3.9329
	social group	0.1574	.4552	1.1705
	constant	−3.6171	.0000	
Girls, 4–10 years				
8	can't concentrate	3.0509	.0241	21.3062
25	doesn't get along	5.5961	.0001	269.3615
85	strange ideas	5.2785	.0030	196.0665
	social group	−0.2849	.4793	0.7521
	constant	−5.3018	.0014	
Girls, 11–16 years				
50	too fearful/anxious	1.5300	.0258	4.6181
85	strange ideas	2.6364	.0129	13.9629
103	depressed	1.6703	.0148	5.3136
	social group	0.1282	.5248	1.1368
	constant	−3.6338	.0000	
All participants, boys and girls 4–16 years				
13	confused	0.8431	.0279	2.3236
25	doesn't get along	1.9804	.0000	7.2454
45	nervous	0.7677	.0268	2.1549
46	nervous movements	1.1193	.0063	3.0626
91	suicidal talk	1.8194	.0001	6.1681
100	trouble sleeping	1.0642	.0030	2.8986
103	depressed	0.8670	.0148	2.3798
	social group	0.0543	.6066	1.0558
	constant	−3.6676	.0000	

*B is the regression coefficient.

45 and item 95: *temper tantrums*. Since *temper tantrums* was prevalent in the normal population (20%), this item was eliminated, leaving three remaining items in the model.

Girls, 4–10 years

Prior to running the analysis, discriminating items with odds ratios less than 20 were excluded (39 items with OR > 10 was too high for the model). In this stratum, a solution of four items remained after running the regression analysis. SES, which in all sub-groups were part of the model, showed a tremendous impact on item 46: *nervous movements*. This encouraged removal of the item from the model. Three items remained, having no interactions. The output of the regression model is seen in Table 18.

Girls, 11–16 years

From Table 17, four items were forwardly selected, although item 89: *suspicious*, was backwardly eliminated, due to the criteria of the model. Amongst the three remaining items, no interactions were present.

Factor analysis

Factor analyses, both principal component and varimax rotated, were performed in four different item matrixes from the clinically based sample. All 118 CBCL problem items and the 89 cross-informant items from the CBCL, TRF and YSR problem section were analyzed. In the 118 CBCL problem item matrix, the first requirement of eigenvalues above 1.0 led to 34 factors. At the first factor, 31 items loaded 0.4 or more. The enlarged inclusion criteria of factor loadings at 0.4 or more, was analog to *Achenbach* (9). The first component was labelled aggressive or externalizing behavior. The second and third factors consisted of 11 items each, all with factor loadings above 0.3. The fourth, fifth, sixth and seventh factors included 10, 7, 7 and 4 items respectively. The remaining factors were all loaded with few and unspecific item combinations.

In the three analyses of the 89 cross-informant items from the CBCL, TRF and YSR respectively, six components materialized. The final structures are presented in Table 19. Selection was based upon item presence at comparable factors in the analyses, of minimum two out of three instruments. Items from the aggressive behavior component (factor load >0.4), which also loaded (>0.3) at another factor, were included at the alternative factor only.

As in Achenbach's original analyses, factors or cross informant syndromes (CIS), were extended with single items taken from the CBCL 118 items factor analysis. Thus, making the final factor structure (as seen in Table 19), more comparable to the US structure. Main externalizing behavior clustered around one single factor, which was called 'aggressive and delinquent behavior'. Three factors were found, that included internalizing behavior. Due to differences in item clustering (compared with Achenbach's study), the factors were labelled different. The labels 'depressed', 'anxious/socially withdrawn' and 'somatic complaints' were used. These three factors were, also, comparable to the US internalizing item pool of 31 items, although the present number of internalizing items amounted to 26 only. No 'withdrawn' factor was exposed, although ten items clustered around a single factor, with the label 'anxious/socially withdrawn'. One factor consisting of six items was named 'attention problems' and four items clustered around a factor that may be analog to the US 7-item 'thought problem' syndrome. As previously mentioned, there were no diagnosed schizophrenias or delusional disordered probands within the clinical sample. Naturally, this influenced the factorial structures, and made it unlikely to counterpart the US cross-informant syndromes.

Latent trait analysis

Exploratory analyses were made on the 118 dichotomized CBCL problem items. Three different sample combinations were analyzed. The CBCL clinical sample (N = 145), the CBCL normative sample (N = 779) and the CBCL mixed sample (N = 924).

One dimensional

Analyzing data in a latent trait model with one dimension, analog to one component, made it possible to determine which items would eventually construct a 'maladjustment scale'. To ensure transferability from a clinical setting to a population setting, only items with factor loadings greater than 0.6 (in all three sample combinations), were included. A total of 24 items (listed in Table 20), fulfilled the criteria. Items were ranked in accordance with symptom prevalence in the normative sample, in order to illustrate increasing severity.

Two dimensional

Whenever factor analysis of the emotional and behavioral problem items on the CBCL and related instruments have been reported, subscales were divided in internalizing and externalizing symptoms or syndromes. From a latent trait model with two

Table 19. Final factor-structure after varimax rotated exploratory factor analysis of 89 cross-informant problem items from the CBCL, TRF and YSR

Externalizing	**Internalizing**	**Neutral**
Factor I: aggressive and delinquent behavior	*Factor II: depressed*	*Factor V: attention problems*
3 argues a lot	12 lonely	1 acts too young
7 bragging	13 confused	8 can't concentrate
16 mean to others	14 cries a lot	10 can't sit still or hyperactive
19 demands attention	18 harms self	11 too dependent
20 destroys own things	25 doesn't get along	41 impulsive, acts without thinking
21 destroys others things	33 feels unloved	64 prefers younger kids
22 disobeys at home*	34 feels persecuted	*Factor VI: thought problems*
23 disobeys at school	35 feels worthless	9 can't get mind off thoughts (obsessions)
25 doesn't get along	38 is teased	66 repeats actions (compulsions)
26 lacks guilt	91 suicidal talk or thoughts	70 sees things
27 jealous	103 unhappy, sad, depressed	80 stares blankly*
37 fighting	112 worries	*The item was added from the CBCL 118 item analysis.
39 bad company	*Factor III: anxious/socially withdrawn*	
43 lying or cheating	31 fears impulses	
48 not liked	45 nervous	
57 attacks people	50 too fearful or anxious	
67 runs away from home*	52 feels too guilty	
68 screams a lot	69 secretive	
74 showing off	71 self-conscious	
86 stubborn	75 shy or timid	
87 moody	89 suspicious	
90 swearing	99 too concerned with neat or clean	
93 talks too much	112 worries	
94 teases a lot	*Factor IV: somatic complaints*	
95 temper tantrums	51 dizzy	
97 threatens people	56a aches, pains	
104 unusually loud	56b headaches	
106 vandalism*	56c nausea, feels sick	
*Three items were added from the CBCL 118 item analysis.	56f stomach aches	

dimensions and varimax rotation (which, in mathematical terms, consists of two axes which are orthogonally angled to each other), the outcome configuration is seen in Table 21. Items with exploratory factor loadings in excess of 0.6 on either dimension in all three sample combinations were included. The two constructs represent 'disruptive' and 'emotional' behavior, and were composed of 16 and 7 items respectively. 'Disruptive' scores between zero and 16, and 'emotional' scores between zero and seven are sufficient for assessment. Mean values, 95 and 98 percentiles of these scores in the population based sample are presented in Table 23.

Multi-dimensional

Four clinically relevant latent traits were generated, these originated from the CBCL clinical sample problem item pool and varimax rotation in five dimensions. All items with exploratory factor loadings in excess of 0.6 ('socially deviant behavior scale' only factor load >0.5) at the attached factor were included. This resulted in a total of 71 included items. To ensure that the structures were more clinically relevant and transferable, some items were excluded, whilst others were included, guided by factor loadings at corresponding latent traits taken from an analysis of the normative sample.

At the first dimension (i.e. factor), thirty-two items associated with the latent trait after exploratory analysis. Three items scanning sexual problems, together with item 3: *argues a lot* and item 78: *smears feces*, were excluded because they loaded extremely differently between settings, with confirmatory factor-loadings below 0.4 in the population based item-pool. The latent trait was labelled 'conduct problem scale'. The items 87: *moody* and 74: *showing off* remained in the model because they loaded homogeneous in both settings and represent mild conduct behavior.

At the second dimension, fourteen items attached to the latent trait after exploratory analysis. The latent trait, labeled 'ADHD problem scale', were reduced by two items, numbered 12: *lonely* and 36: *accident-prone*, in relation to the confirmatory analysis, because these items loaded below 0.6 in both the population based and the clinically based item-pool. Four items, number 19: *demands attention*, 20: *destroys own things*, 25: *doesn't get along* and 104: *unusually loud*, were represented at both the conduct and the ADHD problem scale.

The third dimension was labeled 'depression symptom scale', and included 13 items after

Table 20. Latent trait analysis in one dimension of 118 CBCL problem items. Maladjustment scale — 24 items

Item (short text)		Prevalence* %
87	moody	40.5
41	impulsive, acts without thinking	34.9
19	demands attention	33.9
86	stubborn	31.4
94	teases a lot	27.3
22	disobeys at home	27.1
95	temper tantrums	24.2
34	feels persecuted	20.2
43	lying or cheating	18.5
88	sulks a lot	17.3
38	is teased	13.2
104	unusually loud	10.8
57	attacks people	9.0
37	fighting	7.5
20	destroys own things	7.3
25	doesn't get along	7.0
68	screams a lot	5.6
48	not liked	5.1
13	confused	4.3
16	mean to others	4.1
21	destroys others things	3.1
97	threatens people	2.2
106	vandalism	1.4
18	harms self	1.3

*Prevalence is the reported frequency of the symptom, i.e. scored 1 or 2 by parents in the population based sample.

exploratory analysis. Five items were excluded in the confirmatory analysis (four somatic symptoms and item 32: *needs to be perfect*), all because they were too unspecific. Two items were included (14: *cries a lot* and 91: *suicidal talk*), both because they loaded just below the inclusion criteria in the population based sample, and represent core depression symptoms.

The fourth dimension was less consistent. In the exploratory analysis, fifteen items were included. Two items, 98: *thumbsucking* and 107: *wets during day*, both loaded extremely low in the population based item pool. A number of other items in this latent trait (i.e. the 'socially deviant behavior scale') loaded far from the ideal 0.6 in the confirmatory analysis based on population data, also. Nevertheless, the scale was accepted because of the clinical significance.

The final structures, i.e. scales, are listed with confirmatory factor loadings and ranked according to individual item prevalence in the normal sample in Table 22. Score distributions of the stratified population based sample, at each latent structure scale, are presented in Table 23.

Reliability measures

Interparent reliability

In 45 cases the CBCL was answered by both parents of the referred child. Their instructions were to complete the form without checking each other. Amongst five couples, one parent had more than 10 problem items blank. Fourteen separated couples were excluded, as the parameter of interest was the correlation between adults sharing a common environment with their child. Nine pairs of biological mothers, in parental constellation with stepfathers, were excluded to ensure internationally comparable analyses. The remaining 17 couples of

Table 21. Latent trait analysis in two dimensions of 118 CBCL problem items

Disruptive scale — 16 items

Item (short text)		Prevalence %*
41	impulsive	34.9
19	demands attention	33.9
93	talks too much	33.6
7	bragging	27.7
94	teases a lot	27.3
22	disobeys at home	27.1
10	can't sit still (hyperactive)	21.6
43	lying or cheating	18.5
104	unusually loud	10.8
57	attacks people	9.0
37	fighting	7.5
20	destroys own things	7.3
68	screams a lot	5.6
16	mean to others	4.1
21	destroys others things	3.1
97	threatens people	2.2

*Prevalence is the reported frequency of the symptom, i.e. scored 1 or 2 by parents in the population based sample.

Emotional scale — 7 items

Item (short text)		Prevalence %*
31	fears impulses	34.4
112	worries	23.6
45	nervous	10.5
52	feels too guilty	9.7
50	too fearful or anxious	7.8
103	unhappy, sad, depressed	6.9
18	harms self	1.3

*Prevalence is the reported frequency of the symptom, i.e. scored 1 or 2 by parents in the population based sample.

Table 22. Latent trait analysis — final scales. Factor loads after confirmatory analysis and item prevalence in the the population

Conduct problem scale — 27 items

Item (short text)	Factor loads Populat.	Factor loads Clinical	Prevalence %*
87 moody	0.528	0.549	40.5
74 showing off	0.573	0.575	37.1
90 swearing	0.649	0.726	35.0
19 demands attention	0.692	0.768	33.9
93 talks too much	0.641	0.663	33.6
7 bragging	0.602	0.715	27.7
94 teases a lot	0.780	0.844	27.3
22 disobeys at home	0.794	0.804	27.1
95 temper tantrums	0.715	0.705	24.2
43 lying or cheating	0.662	0.699	18.5
88 sulks a lot	0.624	0.545	17.3
26 lacks guilt	0.481	0.688	17.2
38 is teased	0.538	0.713	13.2
104 unusually loud	0.780	0.849	10.8
23 disobeys at school	0.552	0.721	9.8
57 attacks people	0.718	0.826	9.0
37 fighting	0.713	0.871	7.5
20 destroys own things	0.687	0.715	7.3
25 doesn't get along	0.678	0.763	7.0
68 screams a lot	0.682	0.757	5.6
16 mean to others	0.818	0.883	4.1
21 destroys others things	0.733	0.767	3.1
97 threatens people	0.817	0.859	2.2
81 steals at home	0.484	0.611	1.7
106 vandalism	0.737	0.721	1.4
67 runs away from home	0.614	0.566	1.3
82 steals outside home	0.704	0.629	0.5

ADHD problem scale — 12 items

Item (short text)	Factor loads Populat.	Factor loads Clinical	Prevalence %*
41 impulsive, acts without thinking	0.702	0.770	34.9
19 demands attention	0.633	0.872	33.9
8 can't concentrate	0.726	0.878	26.7
10 can't sit still or hyperactive	0.716	0.813	21.6
64 prefers younger kids	0.585	0.668	20.4
1 acts too young	0.604	0.778	18.4
104 unusually loud	0.710	0.826	10.8
62 clumsy	0.669	0.633	8.5
20 destroys own things	0.701	0.728	7.3
25 doesn't get along	0.704	0.874	7.0
48 not liked	0.749	0.755	5.1
13 confused	0.612	0.680	4.3

Depression symptom scale — 10 items

Item (short text)	Factor loads Populat.	Factor loads Clinical	Prevalence %*
31 fears impulses	0.569	0.445	34.4
112 worries	0.678	0.638	23.6
34 feels persecuted	0.765	0.678	20.2
33 feels unloved	0.849	0.790	15.6
12 lonely	0.627	0.700	13.4
35 feels worthless	0.797	0.777	12.9
52 feels too guilty	0.662	0.651	9.7
14 cries a lot	0.568	0.419	7.4
103 unhappy, sad or depressed	0.858	0.754	6.9
91 suicidal talk or thoughts	0.582	0.544	2.3

Socially deviant behavior scale — 13 items

Item (short text)	Factor loads Populat.	Factor loads Clinical	Prevalence %*
71 self-conscious	0.827	0.658	34.6
17 daydreams	0.354	0.570	33.3
29 fears	0.342	0.570	31.9
75 shy or timid	0.746	0.644	30.9
45 nervous	0.597	0.701	10.5
50 too fearful or anxious	0.685	0.744	7.8
111 withdrawn	0.579	0.687	6.2
80 stares blankly	0.587	0.713	3.6
84 strange behavior	0.395	0.512	3.5
65 refuses to talk	0.490	0.671	3.3
85 strange ideas	0.358	0.642	2.8
66 repeats actions (compulsions)	0.250	0.624	1.9
70 sees things	0.591	0.553	1.9

*Prevalence is the reported frequency of the symptom, i.e. scored 1 or 2 by parents in the population based sample.

biological parents were analyzed. The evaluation of reliability was made by making the sum-difference plots, as seen in Fig. 5, for varying CBCL output (total problem score and the four latent structure scales). A high concordance was found in the majority of couples, although, in some families, parents disagreed considerably. Mothers rated their offspring higher (mean total problem score=46) than fathers (mean total problem score=40). Amongst the four latent structure scales, the poorest inter-parental reliability occurred when a scattered 'socially deviant behavior' appeared in the offspring. Pearson correlation coefficients were calculated and found to be 0.25, only, and therefore non-significant. Mothers scored the depressed and socially deviant behavior in their offspring higher than fathers, who, on the other hand, reported more conduct behavior and ADHD symptoms. The results are discussed and compared with others later.

Test-retest reliability

In 17 referred cases, the same parent answered the CBCL twice, with a mean of 30 days interval. The agreement was expressed by sum-difference plots for total problem score, and for the four latent structure scales in Fig. 6. Once again, a high degree of test-retest resemblance was found within total

Table 23. Mean scores, 95- and 98-percentiles for latent trait scales in the population based sample in age and sex strata

	Boys						Girls					
	4–10 years			11–16 years			4–10 years			11–16 years		
Latent trait scales	M	95	98	M	95	98	M	95	98	M	95	98
CBCL (N = 779):												
Maladjustment scale (24 items)	4.6	14	16	3.5	17	19	3.1	10	12	2.9	10	13
Disruptive scale (16 items)	3.8	10	12	2.8	11	13	2.5	8	10	1.8	6	8
Emotional scale (7 items)	0.8	3	6	1.1	4	5	0.8	3	5	1.2	4	6
Conduct problem scale (27 items)	4.8	14	15	4.1	15	18	3.3	11	13	2.9	10	12
ADHD problem scale (12 items)	2.7	8	9	2.4	8	11	1.5	5	6	1.5	6	7
Depression symptom scale (10 items)	1.4	6	7	1.6	7	8	1.3	5	7	1.6	6	8
Socially deviant behavior scale (13 items)	1.7	5	7	1.6	4	7	1.7	5	6	1.7	5	6
TRF (N = 547):M												
Conduct disorder scale (24 items)	3.3	15	19	2.9	11	15	1.2	7	11	2.0	10	11
ADHD scale (12 items)	2.4	9	10	2.5	9	10	1.0	5	8	1.6	7	8
Depression scale (10 items)	1.3	6	7	0.9	5	6	1.1	5	7	1.3	5	7
Socially deviant scale (13 items)	1.4	5	6	1.1	4	6	1.4	5	7	1.3	5	6
YSR (N = 355):												
Conduct disorder scale (26 items)				6.5	16	18				5.1	11	13
ADHD scale (12 items)				3.3	9	10				3.0	7	9
Depression scale (10 items)				2.0	7	8				2.7	8	9
Socially deviant scale (12 items)				2.7	8	10				3.6	8	9

*The 95%-ile is the recommended cutoff in screening and the 98%-ile represents the definite clinical range.

problem score, conduct problem scale score and ADHD problem scale score. More test-retest deviation was found in the depression symptom scale score, and variance was most evident in the socially deviant behavior scale (Pearson corr. was 0.58 and $P=0.02$). In two or three cases, the difference between test and retest total problem score was enormous, with retest scores between 20 and 50 points higher than the first test score. In the sum-difference plot for the depression symptom scale, one child regressed from four depressive symptoms in the first test, to no symptoms at all in retest. A further child had a reported progression from zero depressive symptoms in the test, to seven in the retest. Some of these huge differences may be explained by actual change in psychopathological status during the 30 days between the two tests, and it is, therefore, not a characteristic connected with the psychometric instrument.

Cross-informant reliability

In the inter-informer reliability analysis, considerably more data was available, thus, both the population based sample and clinically based sample were evaluated. To avoid Pearson correlation, simple two-by-two table analysis was used as the fundamental for calculating the odds ratio, that a child was rated in the clinical range by ex. the teacher, if already so by his parent. The clinical range was defined by having scores equal to, or above the 95%-ile derived from the Danish norms. The results are listed in Table 24.

With the exception of rating ADHD symptoms, there was, generally, better consensus between parents and youths, compared with that of parents and teachers. Socially deviant behavior was rated differently by the three informers, with odds ratios 5.3 and 7.4 for parent/teachers and parents/youths respectively.

Odds ratios for youths rating themselves in the clinical range, given that their teacher had previously done so, were, in all outputs, significantly lower than the comparisons mentioned in Table 24. Although it was apparent that parents were the best informers, in the sense of being positive discriminators between referred and general-population children, no further analyses were made to compare teachers' and youths' scores. When comparing informers rating in the clinically based sample separately, parent/teacher correlation was considerably higher than parent/youth correlation.

EXTERNAL VALIDITY

Screening abilities

In order to evaluate the sensitivity of US scoring profiles, the 95%-ile cutoff was used in accordance with the manuals (9, 14, 15). Being a referred child was the golden standard, and a screen-positive proband was defined as an individual with scores equal to, or above cutoff, at either of the following eleven output scores: total behavior problem score, internalizing problem score, externalizing problem score or one of the eight cross-informant syndromes.

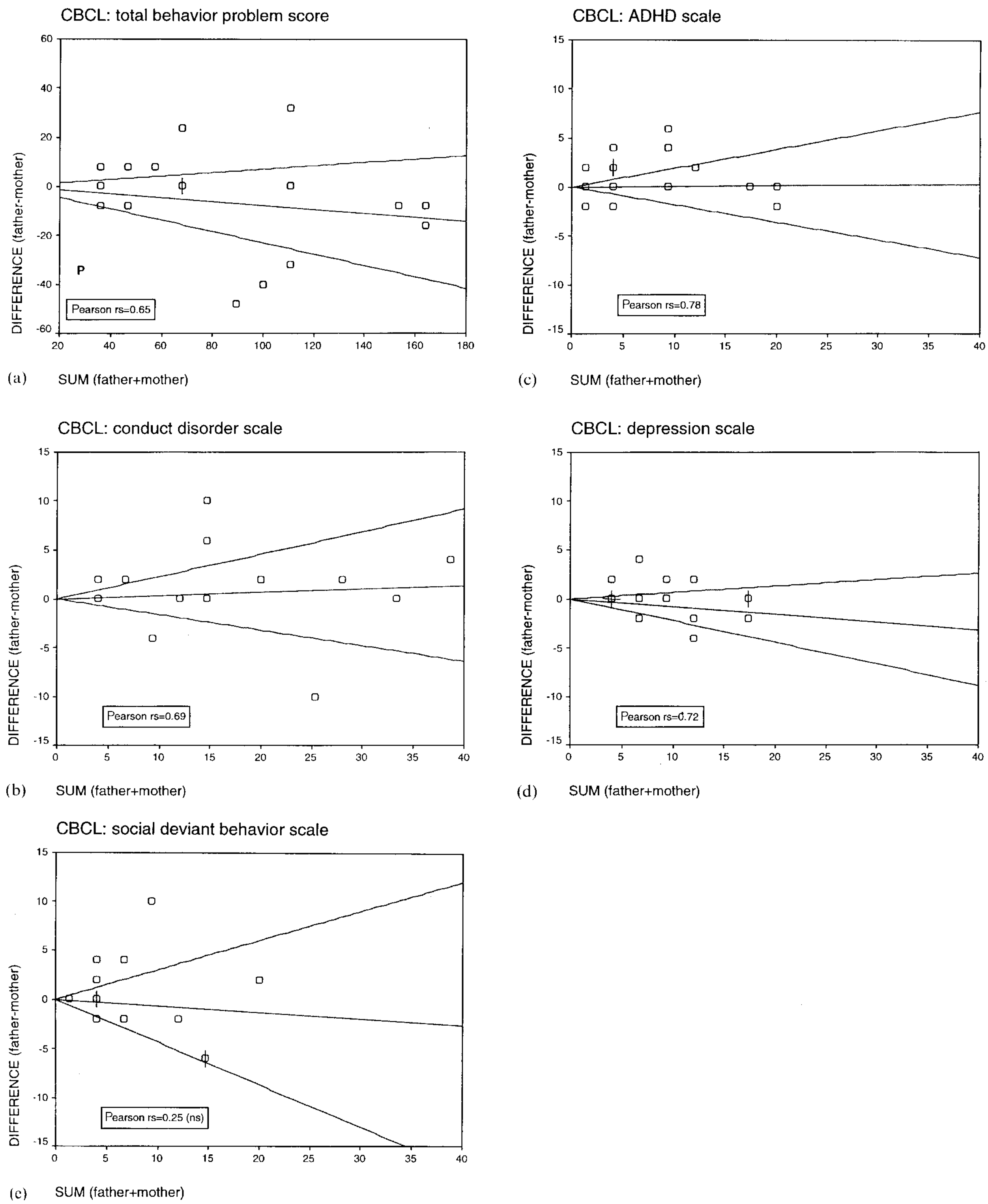


Fig. 5. Inter-parent reliability. Sum-difference scatter-plots for CBCL total behavior problem score and the four latent trait scales. Annotation of Pearson correlation coefficients (n = 17).

Supplementum

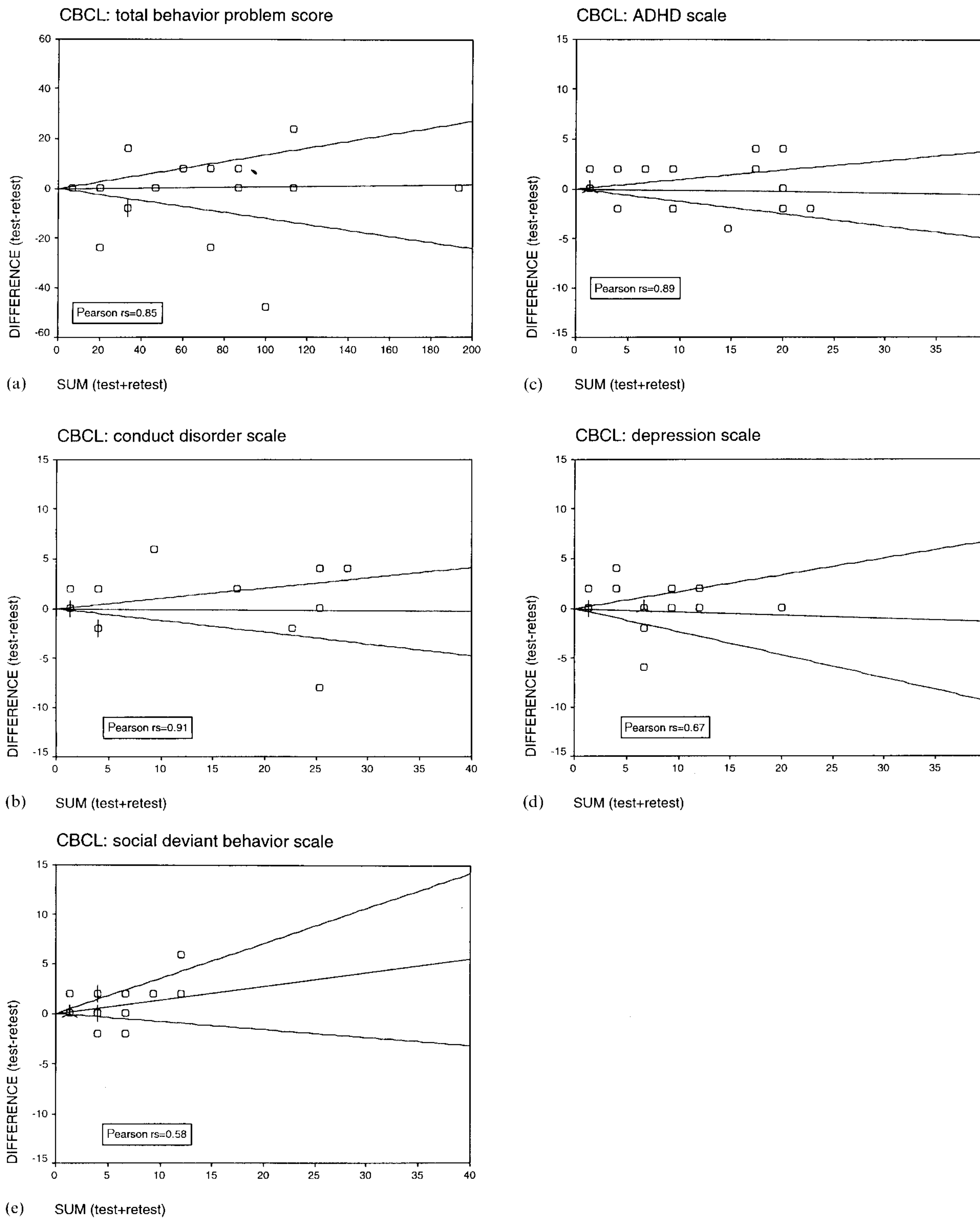


Fig. 6. Test-retest reliability. Sum-difference scatter-plots for CBCL total behavior problem score and the four latent trait scales. Annotation of Pearson correlation coefficients (n = 17).

Table 24. Inter-informer reliability

	Odds ratios, that teachers and youths also score in clinical range*	
CBCL output in the clinical range*	TRF OR	YSR OR
total problem score	12.5	14.7
A) conduct behavior scale	12.5	18.5
B) ADHD scale	16.0	13.5
C) depression scale	8.7	15.7
D) social deviant behavior scale	5.3	7.4

*Clinical range was defined as scoring above the 95%-ile derived from the Danish population norms (Fig. 3 and Table 23).

This algorithm was applied to the clinical sample and sensitivity was measured.

Sensitivity of the CBCL

The ability of the CBCL, classifying the cases correctly was calculated to 0.79. Twenty percent or 155 out of the 779 general-population children, were test-positive due to the algorithm. Specificity is not introduced here, as in this study there was no access to information regarding present but unrecognized pathology in the population-sample. Following the recommended 95%-ile cutoff from the US 1991-profile (9), total behavioral problem score from the CBCL isolate, showed a sensitivity of 0.70. Less than 50% of the 28 responding referred girls, 11–16 years, were screen-positive regarding this specific output.

Sensitivity of the TRF

When testing the same algorithm at the output from the TRF emotional and behavioral problem section (14), the sensitivity was 0.78. It is significant that 17 (89%) of the 19 teacher-rated referred girls, 11–16 years, were identified as test-positive.

Sensitivity of the YSR

The YSR identified slightly more than 50% of the cases only, using the US profile (15).

Sensitivity of combined use

A combination of the CBCL, TRF and YSR output scores provided high sensitivity. A total of 142 (or 90%) of all participating referred children (N=157) were screen-positive. Specificity would probably be low, suggesting that, used as a screening instrument in a population study, one must expect an increased number of false-positive probands. A more valid evaluation of sensitivity, specificity, predictive value of positive test and screening cost-benefit must wait on a two-phase epidemiologic study in progress.

Diagnostic use

Diagnostic value of the CBCL (US cross-informant syndromes)

In order to evaluate the diagnostic value of the CBCL more specifically, the US cross-informant profiles and cutoffs were used once more. The number of clinically assessed responders diagnosed within the following ICD-10 categories; OCD (F42), adjustment disorders (F43), eating disorders (F50), pervasive developmental disorders (F84), conduct disorder (F91), emotional disorders (F93) and disorders of social functioning (F94), were compared with the number of cases identified by the different syndrome scores. Both main and comorbid diagnoses were included, and the 95 percentiles were used as cutoff. The results can be seen in Table 25.

Eating disorders were somewhat silent and would frequently fail to be identified in a screening survey, using standard outcome from the CBCL as first step. The majority of patients in other diagnostic

Table 25. ICD-10 diagnosed children within certain diagnostic categories and the corresponding number screened by each CBCL scale (US-profiles)

	Number of patients identified by different CBCL scales**									
ICD-10	Internalizing				Neutral			Externalizing		
clinical diagnosis* (n)	1	2	3	In	4	5	6	7	8	Ex
F42 Obsessive-compulsive disorder (5)	0	1	0	2	0	4	0	0	0	0
F43 Reaction to severe stress, adjustment disorder (10)	3	1	5	6	5	4	5	4	3	5
F50 Eating disorder (8)	2	2	2	1	1	2	1	1	0	1
F84 Pervasive developmental disorder (24)	9	7	11	15	12	7	8	7	5	12
F90 Hyperkinetic disorder (4)	0	0	0	0	4	1	4	4	4	4
F91 Conduct disorder (30)	6	8	8	15	21	6	14	15	12	16
F93 Emotional disorder (14)	2	3	7	7	6	4	2	1	1	1
F94 Disorder of social functioning (7)	1	2	3	3	5	3	4	2	3	4

*Both main and comorbid diagnoses were included from 126 assessed children.
**Screened by scale scores above the 95%-ile (borderline clinical range) due to US standards.
Scales are: 1: withdrawn; 2: somatic complaints; 3: anxious/ depressed; In: internalizing; 4: social problems; 5: thought problems; 6: attention problems; 7: delinquent behavior; 8: aggressive behavior; Ex: externalizing.

groups would, presumably, be identified, although the precision by which standard outcome from the CBCL distinguished diagnostic categories was poor. No more than 50% of the probands with pervasive developmental disorder scored in the clinical range at the 'social problems' scale. Only 16 of the 30 referred children diagnosed with conduct disorder scored above cutoff at externalizing behavior. Conduct disordered children were, actually, screened more satisfactorily by the 'social problems' scale, where 21 of the 30 children scored above the 95 percentiles. Greater consistency appeared to follow disorder of social functioning and the 'social problems' scale, as well as hyperkinetic disorder and OCD, respectively, 'attention problems' scale and 'thought problems' scale. As anticipated, the symptoms of probands with adjustment disorder and emotional disorder were spread over a number of different scales.

Diagnostic trend in the latent trait scales

The latent trait scales (Table 22) were not tested as diagnostic predictors. Such analyses would be inappropiate, though the scales were generated from the same material. Evaluation of the validity of these scales must attend new clinical data. Nevertheless, different diagnostic categories within ICD-10 were evaluated by their score-levels at the four CBCL latent trait scales (Table 26). Once again, eating disorders were somewhat silent, although these patients (all girls) scored high on the 'socially deviant behavior scale'. Children with adjustment disorder and disorder of social functioning had similar scoring profiles, with high or borderline-high scores at all four latent trait scales. Hyperkinetic and conduct disordered patients, who in clinical picture often look alike, seemed possible to separate because hyperkinetic children, in contrast to conduct disordered children, were rated below mean of baseline population on the 'socially deviant behavior scale'. Children with conduct disorder, in addition, scored high on the 'depression symptom scale'.

DISCUSSION

Population based epidemiology within child and adolescent psychiatry would be impossible without the use of validated psychometric instruments (i.e. those which have the predictive power to screen larger samples). The aim of this study was to standardize, cross-national compare and refine the CBCL and related material by Achenbach and Edelbrock. In this discussion, various methodological approaches will be systematically debated. The most interesting results will be reviewed and compared with analogous outcome known from the literature. Finally, the appropriate follow up studies will be outlined.

Design

Sampling

Many different ways of sampling have been used in order to achieve national norms within the CBCL literature. This fact makes it difficult to perform valid comparisons (26). There is some scepticism as to the matching of a normative sample similar to that of Achenbach in the 1991 US profile (9), which is likely to camouflage significant differences between child psychiatric patients and the general population. If, for example, family SES, or single-parent-families were significant predictors for referral, one may overlook valuable information when comparing matched samples. Sampling procedures including selected school-based samples (16, 17, 20, 31, 36, 72–74), have the risk of being super-samples. Public schools, in general, are not representative. Children from the higher social-classes e.g. attending private schools, or children who receive special instruction may be under-represented in such a sample. In the French study, Fombonne (16, 72), compensated for this by mixing public and private schools and by oversampling children attending special classes. The first Greek sample (17) was a public school sample taken from the Greater Athens area only. In this sample, extremely high mean scores were found. In a later study, Roussos et al. achieved a more representative sample, covering both urban and rural areas of Greece, and taken from both private and public schools. The mean problem scores in the latter were more comparable to those from other cultures (75). The Swedish sample (31) was taken from public schools in Uppsala primarily, a larger city north of Stockholm. The Swedish group has, along with the German group, presented the lowest total problem scores from larger community samples ever published, and both samples are suspected of being non-representative.

Exclusion criteria, also, makes results incomparable, owing to the use of varying strategies. The majority of authors have excluded children receiving psychological treatment and special education (9, 31, 72), in order to create normal samples, in contrast to referred samples. These exclusion procedures follow the recommendation of Achenbach, when making cross-cultural comparisons and when making discrimination analyses in accordance with the ANOVA design (in which a referred sample is compared with a normal or non-referred sample). Nevertheless, exclusion criteria reinforce improved functioning and less behavioral pathology amongst children in the national normative samples, compared with unselected population based samples. In their first Dutch sample published in 1985 (1), Verhulst et al. made a clear distinction between the normal sample and the population based sample.

In the present study, an epidemiological fundament was preferred. The population based sample represents a base in a 'case-base study design', rather than a non-case sample. No children were excluded from the population based sample, despite referral to child psychiatric services, child guidance clinics or identified as being deviant in any way. This approach encourages the tendency to use normative outputs as reference data for typical case/control designs. A non-limited and ideal representative population based sample encompasses the most valuable information, when the main interest is exactly what separates a deviant child from the average child. By using a random sample of sufficient numbers, taken from the population register, there was a guaranteed representation of onset. The 'golden standard' applied in order to select cases was 'referral to child psychiatric services'. This sampling procedure encourages recall-bias, because the person who fills in the form, is aware of proband status. The current design, also, restricts analysis of external validity.

Data collection

The method in which the data were collected became a further challenge, with regard to comparison with other studies, as well as in the struggle of undertaking the methodologically right thing. The basic assumption was that data should be collected in a way parallel to the way in which the instruments would be used later. Consequently, the mailing procedure was chosen. Achenbach in the US (9), and Verhulst in The Netherlands (1), both collected normative data by interviewing parents within their homes. It is impossible to determine in which direction this procedure could have biased their results. In a number of studies (in which school-based samples were used), the checklists were distributed to parents via the classroom teacher. The French, the Swedish, the Icelandic (partially) as well as a number of other normative samples were collected this way. This procedure appears to have diminished attrition, probably owing to the fact that parents have felt more responsible. Mailing questionnaires to parents in a stratified random community sample (as in this study) was, also, conducted by the German group, and in the Norwegian study (34, 39, 76).

Representation

The best possible way to analyze representation of an incomplete population based sample in epidemiology, is to force some non-responders to answer the questionnaire. Using telephone contact and persuading a small random sample of non-responding families to participate would have strengthened the results of this study. This design was planned, but not permitted by the Danish ethical committee. Nevertheless, comparable demographic data was obtained from the Danish Statistical Institute relating to both the responders and the non-responders, thus generating a number of statistical contrasts.

Although the response rate of 0.61 was low, compared with the American and Dutch normative sample (respectively 0.90 and 0.80), the rate was reasonable when correlated to other mailed surveys. In Norway, Növik reached a response rate of 0.45 and in the first German normative sample the participation rate was 0.55 (34, 39). The attrition analyses made on register data showed that the normative sample was under-represented by single-parent families and the lowest family SES group. Additional results demonstrated; 1) significant higher mean total problem scores in lower SES children at the CBCL and TRF, 2) boys of single mothers scored higher than boys from core families, and 3) all competence and problem items marked with an 'S' in appendices A.1, A.2, A.3 and B were significantly more deviant scored in lower SES children. This bias should be considered when interpreting the results. Prevalence proportions and scores were, probably, somewhat underestimated, whereas discriminative power would be a little overestimated.

Total behavior problem scores

Total behavior problem scores, externalizing and internalizing problem scores derived from the CBCL and TRF (at the YSR to a much lesser extent) were documented, as in all the CBCL literature, to differentiate between children in the population and referred children at sample level. *T*-test for equality of means with unequal variance was used and highly significant differences were found. The boxplots in Fig. 4. illustrates the same. The present study was not designed to estimate the best cutoff, although it seems appropriate to choose a cutoff in total problem score at the CBCL around 30 (Fig. 4).

Significant sex difference was found in total behavior problem score, with boys scoring higher than girls at the CBCL and TRF. The difference was most apparent amongst younger children and diminished amongst older age-groups. This result was parallel to the findings of a 12 culture comparison by Crijnen et al. (26). No significant age differences in total behavior problem score were found in the population.

The mean total behavior problem score for the Danish population based sample, CBCL/4–16 years, was 17.7 (s.d.14.9), which was somewhat low compared with most other cultures. In Fig. 7, the mean total problem scores from normative data in different cultures are illustrated (beware of the variations in age-span). The mean values were found in original papers, or in cross-cultural reviews.

The close resemblance between the Scandinavian norms are amazing. The Swedish total scores are lower than the Danish, Norwegian and Icelandic, discussed by Larsson et al. (31).

Item analysis

Statistical approaches

Differences in statistical approach, logistic regression analysis *vs.* analysis of variance, complicate the comparison of 'single item discriminative power measures' generated in this study (odds ratios), and 'effects of referral status' (percent of variance accounted for by single items), which has been conducted in most other studies. The argument for dichotomizing item response and using logistic regression analysis was, mainly, to avoid the ANOVA design. The latter method demands that item score follows a normal distribution, *which* is certainly not the case. From appendix B, it was obvious that many item-scores, primarily in the clinical sample, were distributed somewhat asymmetrically. e.g. item 41: *impulsive* in 4–10 years referred girls dispersed, 47 percent scored '0', six percent scored '1' and 47 percent scored '2'. By dichotomizing item scores with cutoff between 0 and 1, information was reduced and the risk of systematic error may have increased. As a pilot analysis, different cutoff levels were chosen. In prevalent symptoms such as item 3: *argues a lot*, 7: *bragging*, 27: *easily jealous*, 29: *fears*, 31: *fears impulses*, 32: *needs to be perfect*, 41: *impulsive*, 71: *self-conscious*, 74: *showing off*, 90: *swearing* and 93: *talks too much*, discriminative power and predictive value would have increased by choosing a cutoff between 1 and 2 (without losing too much statistical strength). On the other hand, the decision was made to generalize the problem item cutoff by practical means. When re-examining the scoring profiles and the latent structure analysis, this will be one of the subjects to reconsider.

Competence items

Competence items at the CBCL, TRF and YSR are of most inhomogeneous value. Due to great consensus regarding limited validity of the activity scale items (page one in the CBCL and YSR) (51, 64), the decision to exclude these items from analysis was made. There was valuable information in the social competence items. Item V.1/CBCL: *number of good friends* and VI.b/CBCL: *behavior with others* were extremely potent discriminators between population and referred children. The scattering of competences supports the fact that social-functioning at home, in school, or in other contexts were main predictors of psychological and psychopathological outcome. If a child socializes well, he or she was unlikely to be emotional and behavioral deviant. Academic performance was less significant as a predictor, although as a group, referred children were generally poor achievers with regard to reading, spelling and mathematics. This result is consistent with the outcome of other studies (1, 9, 72).

Problem items

When comparing the present analyses of behavioral and emotional problem items with other studies including well documented discriminative analysis of CBCL-, TRF- and YSR- problem items (1, 9, 10, 16–23, 36, 72, 74, 77, 78), a number of similarities

Table 26. ICD-10 diagnosed children within certain diagnostic categories and the corresponding score-levels at latent trait scales on the CBCL

	Score-level at latent trait scale			
ICD-10 clinical diagnosis* (n)	Conduct	ADHD	Depression	Soc. deviant
F42 Obsessive-compulsive disorder (5)	—	—	↑↑	↑↑
F43 Reaction to severe stress, adjustment disorder (10)	↑	↑	↑↑	↑↑
F50 Eating disorder (8)	—	—	—	↑↑
F84 Pervasive developmental disorder (24)	↑	↑	↑	↑↑
F90 Hyperkinetic disorder (4)	↑↑↑	↑↑↑	—	—
F91 Conduct disorder (30)	↑↑	↑↑	↑↑	—
F93 Emotional disorder (14)	—	↑	↑	↑↑
F94 Disorder of social functioning (7)	↑	↑↑	↑↑	↑↑

*Both main and comorbid diagnoses were included from 126 assessed children.
'-' = low score or average score; '↑' = borderline high score; '↑↑' = high score; '↑↑↑' = very high score.

occur. The CBCL items listed in Table 11 materialized as excellent discriminators, with odds ratios above 10 ($P=0.01$). Between cultures, there appears to be consistency about items 103: *unhappy, sad or depressed*, 13: *confused*, 25: *doesn't get along* and 45: *nervous* as superior discriminators. Opposite, item *2: allergy* and *4: asthma*, are generally isolated as poor discriminators. Eight of the 12 CBCL problem items, considered to be poor discriminators in a Dutch study (1), reappeared as being non-discriminators in Table 10 (item 4, 5, 28, 32, 77, 99, 105 and 110). In a French study (72), thirty problem items failed to discriminate, eleven of these were contained in Table 10. In the largest study of 4,220 demographically matched referred and non-referred children (9), Achenbach found that items 2: *allergy* and 4: *asthma*, together with item 75: *shy or timid*, 83: *stores up unneeded things* and 99: *too concerned with neat or clean* did not discriminate with a significance level of $P<0.01$.

In a number of symptoms, sex differences appear to be somewhat robust between cultures. Patterns similar to those of other cultures, emerged in the Danish population based sample. Boys, generally, exhibited more externalizing behavior than girls. Eleven CBCL problem items with sex differences in five out of six comparisons (16, 18, 21, 73, 74, 77) scored higher in boys in this study also. The items were number 7: *bragging*, 8: *can't concentrate*, 10: *hyperactive*, 23: *disobeys at school*, 37: *fighting*, 41: *impulsive*, 61: *poor school work**, 72: *sets fire**, 74: *shows off*, 94: *teases a lot* and 95: *temper tantrums*. With the exception of two items (marked with an *), all others were either, in the 'conduct problem scale', or the 'ADHD problem scale' found in the latent trait analysis. Three comparisons (22, 73, 78) provided a list of ten TRF problem items, all scored significantly higher in boys than in girls by classroom teachers. Nine of these items; 2: *hums, makes noise*, 4: *fails to finish things*, 10: *hyperactive*, 15: *fidgets*, 37: *fighting*, 41: *impulsive*, 61: *poor school work*, 62: *clumsy* and 78: *inattentive* showed the same significance level in the present study.

Age differences were common at item level, although small or non-significant at total problem score level. In four comparisons (18, 21, 74, 77) age effects were found for seven CBCL problem items, all indicating higher scores in the case of younger children. In five of these items, number 19: *demands attention*, 20: *destroys own things*, 24: *doesn't eat well*, 29: *fears* and 108: *wets bed*, the Danish study corresponded. In this study, twenty CBCL problem items showed significant age effect, with higher scores in older children. Only one item, 61: *poor school work*, reappears in all studies with reported item analyses. Some interpretational difficulties must be assumed, when a checklist is expected to cover an age interval of four to 16 or 18 years. As a result of this, scoring must be conducted both age and sex specific, and referenced to stratified normative data.

SES differences were found in the present study on total problem score level at the CBCL and TRF, but not at YSR. Compared with the statistical strength of referral status, SES showed a lesser degree of implication, although once again, SES accounted for more statistical variance than gender effects. Findings were analog to other studies (10). At item level, no single problem item was scored significantly higher amongst upper SES children, compared with a significant SES influence in 33 CBCL problem items, with lower SES group scoring higher. International comparisons agree in a number of CBCL problem items with SES differences in the latter direction (10). Items 8: *can't concentrate*, 10: *hyperactive*, 11: *too dependent*, 37: *fighting*, 38: *is teased*, 43: *lying or cheating*, 53: *eats too much*, 90: *swearing*, 93: *talks too much* and 99: *too concerned with neat or clean*, reappear in the majority of studies, of which all but item 99 showed the same significant difference in this study also. The consensus in other studies regarding SES effects on teachers reporting of problems was not parallel to the present findings. Danish teachers tended to register less SES influenced problem behavior.

Cross-informant correlations

When teachers rated items having an element of comparison with other children of the same age, some items became improved discriminators, or as good as parents rating (e.g. items 28: *eats nonfood*, 44: *bites fingernails*, 55: *overweight*, 63: *prefers older kids*, 96: *thinks about sex too much*, 99: *too concerned with neat or clean* were better discriminators in teachers scattering). On the other hand, when scoring involved more intimate knowledge relating to the child, teachers scored both referred and background population lower. Typically, TRF in contrast to the CBCL, did not differentiate groups on items 56a-g: *somatic complaints*, 40: *hear things*, 70: *see things that aren't there*, 52: *feels too guilty* and 30: *fears school.*

The short screening constructs

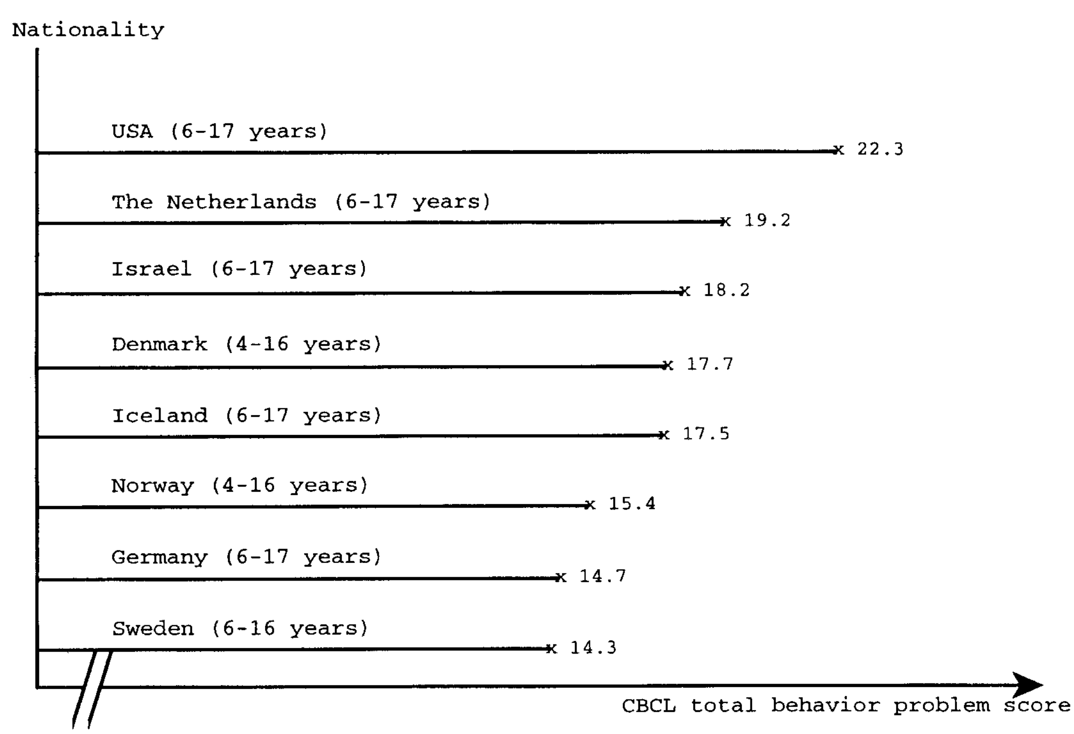
The alternative screening constructs derived from the multivariate regression analyses in age and sex strata may have interesting prospects. If, in a proper external validation, the different item combinations from Table 18 turns out highly predictive, only two or three specific symptoms reported present by parents might be sufficient as case predictors. Such small tools might be extremely useful in primary care and school medicine. The value of the constructs will be evaluated in a two-phase study in progress.

Comparisons of factorial structures

Principal component analysis

Exploratory factor analysis, or principal component analysis followed by varimax or promax rotation, has been part of an internal validity examination in a number of studies (9, 24, 25, 38, 39, 67, 79). Table 19 represents the result of a similar factor analysis made on Danish referred children. The analyses contained only 146 CBCL forms, 118 TRF forms and 58 YSR forms, which are few and, possibly, insufficient for valid comparison purposes. The 'aggressive and delinquent behavior' factor was most similar to the total externalizing problem item pool found by Achenbach in his original factor analysis (9). However, in the Danish clinically based sample it was impossible to distinguish two externalizing syndromes, 'aggressive behavior' and 'delinquent behavior'. Out of the 28 items in the united factor, 26 items were represented in either of Achenbach's two externalizing cross-informant syndromes. Items 25: *doesn't get along* and 48: *not liked*, clustered amongst externalizing items. In the US factor analysis, both were found at a 'social problem' factor. On the other hand, seven items from the US 'delinquent behavior' factor, item 63: *prefers older kids*, 72: *sets fire*, 81: *steals at home*, 82: *steals outside home*, 96: *thinks about sex too much*, 101: *truancy* and 105: *takes alcohol or drugs*, all failed to appear at any of the factors in this study. Findings were, also, parallel to a Dutch and French comparison (67), where problem items loading at the same factor in both cultures (clinical samples) were identified. All 12 aggressive behavior items in their combined study, also, materialized in this factor analysis.

A reduced level of consistency is found when examining internalizing and neutral factors. A 'depressed' factor either materializes solely, or fuses with 'anxious' or 'withdrawn' factors in various studies. Agreement appears to occur relating to items 12 *lonely*, 33: *feels unloved*, 34: *feels persecuted*, 35: *feels worthless* and 112: *worries*, attaching to a 'depressed' factor. Items 91: *suicidal talk* and 103: *unhappy, sad or depressed*, also, connect in the majority of factor analyses. No other studies, other than the present, have found items 13: *confused*, 25: *doesn't get along* and 38: *is teased*, attached to the 'depressed' factor. By appearing in both externalizing and internalizing structures, item 25: *doesn't get along* becomes a particularly potent predictor. Somatic symptoms were expected to cluster. Items 56d: *eye problems* and 56e: *skin problems* fail in connecting to the 'somatic com-



Note: Results were gathered from different sources.

Fig. 7. Mean total behavior problem score at the CBCL. Comparison of different national normative samples.

plaints' factor in this study, as in all other studies. Regarding the 'attention problem' factor, the majority of authors would appear to agree in items 1: *acts too young*, 8: *can't concentrate*, 10: *hyperactive* and 41: *impulsive* being attached. Unexpectedly, items 61: *poor school work* and 62: *clumsy* did not connect with the other ADHD core symptoms.

Latent trait analysis

The process, in which the latent trait or structures was generated, was different to the process from which the standard factors or syndromes are generated. Standard factor analysis is made on continuous variable output (0–1–2). In contrast, the latent trait scales in this study were generated from dichotomized variables. Instead of conducting an analysis of data collected from referred children only (which has been standard in other studies), the analyses were made from the data of referred children and the population based sample separately, and in combination. Exploratory analysis of all problem items in one dimension resulted in a 24 item 'maladjustment scale' (Table 20). It was clear, after examination of the list, that items reflected mixed, predominantly externalizing behavior following child psychiatric illness, as opposed to core psychopathologic symptoms. Due to the wide spectrum of child psychiatric symptomatology, there is no logic in working with a common psychopathology scale, constructing subscales is more logical.

In a two-dimensional model, the factor loadings of attached items at each dimension, following exploratory analysis, were considerably more homogeneous than in the one-dimensional model. The prevalence in the population of connected items disperse over a wide interval. This indicated that the 'emotional' and 'disruptive' traits (Table 21) were approaching scales.

The multi-dimensional model resulted in four clinically relevant scales. Confirmatory factor loadings from analyses of both the population based and the clinically based data are shown in Table 22. With the exception of factor loadings derived from the population based sample, at the 'socially deviant behavior scale', there was homogeneity in loadings at the various scales. Items were rather equally correlated to the latent traits (i.e. equals parallel slopes at normal ogive curves in Fig. 2). Consequently, each symptom attached to a latent trait is part of a thing that exists as a phenomenon, both in a clinical setting as well as in the population. This is referred to as transferability between different contexts. Perhaps a depression is two different things, with one manifestation in a clinical setting, and another in a community setting. If, as in epidemiology, incident cases in the population are the main interest, it is essential to use an instrument calibrated correctly for assessing e.g. depressions in the community.

Reliability measures

The study was designed to analyze inter-parent and test-retest reliability on data from the clinically based sample. This decision made it impossible to measure correlation by odds ratios of parents' agreement in scoring the offsprings in the clinical range, in different outputs from the CBCL. Almost all probands would be expected to receive high scores from both parents, and to be scored high both the first and the second time, in the test-retest design. Alternatively, reliability was tested by studying sum-difference plots, including confidence lines (Figs 5 & 6). Unfortunately, these plots cannot be compared to standard reliability measures. The most used reliability tests, Pearson correlation coefficients, were calculated for comparative use.

Interparent reliability

Interparent reliability measured by Pearson correlation was estimated at 0.76 by Achenbach (9), compared with 0.65 in this study. In the US, all differences between parents reflected higher scores in the mothers' rating, in contrast to the Danish results, where fathers report more externalizing behavior in referred children than mothers.

Test-retest reliability

Test-retest reliability was 0.93 (one week test-retest interval) in the US study (9), and 0.85 in this study (30 days test-retest interval). Comparisons are difficult to interpret due to varying test-retest intervals and, as the analyses in this study, were based on referred children.

Crossinformant reliability

In the majority of studies parents score their offspring higher than a teacher would score the child (80). It has, also, been confirmed by many authors that parents and teachers have a higher correlation, when rating externalizing, as opposed to internalizing, behavior. In larger studies, some authors have tried to evaluate agreement between parents' reports and adolescents self-reports regarding problem behavior (81, 82). As in this study, higher correlation is found between parents and their teenage offspring, than between parents and teachers. Youths, generally, report more problems themselves than their parents. Discrepancies are increased for externalizing rather than for internalizing problems, and increased for girls rather than for boys. As in this study, older girls, in particular, are at risk of having and reporting problems that others fail to recognize. The findings indicate that adolescents are indispensable informants with regard to their own behavioral and emotional problems.

External validity

Applying all scales and total scores of the three questionnaires according to the US manuals, 90% of all referred cases were screened positive. This was acceptable, although, specificity would probably be low, creating many false positive cases who would be incorrectly screened. Children with the most common ICD-10 psychiatric diagnoses, except from eating disorders, seemed to be fairly well identified by joint use of the CBCL materials.

External validity, including predictive power, of the questionnaires in a Danish population must await a further study, which will be performed in the near future. The latent trait scales, and the short screening constructs, need to be evaluated in new samples also.

FINAL CONCLUSIONS

The primary aim of this study was to introduce a validated screening and psychometric instrument within Danish Child and Adolescent Psychiatry. A number of instruments were available, and the decision was made to choose a questionnaire, that was both well known and well described, due to the obvious benefits of using, comparing and publishing research. Although statistical approaches that differed from the ones previously used, it would seem reasonable to compare the results from this study with similar studies performed amongst other cultures.

1) The CBCL and related material, provides good construct validity. Social competence items, the majority of school competence items and all but 19 out of the 118 problem items at the CBCL, discriminated significantly between referred and population based samples. The combination of items within the instrument seem to reflect the phenomenons of interest with substantial precision.

Regarding the content validity, most child and adolescent psychopathology, except from eating disorders, is sufficiently covered by the combination of items.

Factor analyses contribute with important information about the construct and content validity. The eight cross informant syndromes, introduced by Achenbach, have generally been under critique of being somewhat inconsistent across cultures and of little diagnostic and clinical relevance. This was confirmed by the present study. The latent trait analyses provided us with four new scales, which fulfill a number of criterias that makes them promising. These scales must, nevertheless, be thoroughly validated in new studies before they can be generally released.

Item bias, i.e. problems interpretating data equally, is inevitable, when an instrument claims to cover symptoms in both genders and in an age interval from four to 16 years. It is essential, always, to collect information about family social status (SES), to be able to compare data from individuals or samples.

2) In this study, the robust concept of multiple informants has, also, proved a necessity in ensuring high validity. Teachers were evaluated to be aware of problem behavior in referred children, as the majority of children referred to child psychiatric services are actually identified jointly by schools and parents. Children with behavioral problems or emotional disorders not identified by teachers are more frequently false negative cases. These children, often internalizing, must be identified by setting more emphasis upon education and the guidance of parents and professionals working in the social or educational system.

3) Reliability, both inter-parent and test-retest, was acceptable. Although, mothers seem more aware of internalizing behavior, whereas fathers report more externalizing behavior in their offspring.

4) The refinement of the instruments, due to the logistic regression models, opens-up new aspects of interpretation. The short screening constructs may be useful for professionals in the primary health care section (e.g. school health-care system), but their predictive value must be undertaken in new studies.

Recommendations

The CBCL and related materials should now be used as routine procedure in clinical settings and for various research purposes. Epidemiological studies of different subsamples of children at risk is evident use of the questionnaires and outputs can be compared with the reference population or other samples of interest. It is recommended to use all three informant versions and always to collect data about family socio economical status (SES). Total behavior problem scores, externalizing and internalizing problem scores from each instrument are internationally widespread and valid measures. Pay attention to the fact that the eight cross informant syndromes are not sufficiently valid and transferable between cultures and therefore must be used with care. Single item analyses can be very informative and especially competence items, *number of friends* (CBCL;V1) and *behavior with others* (CBCL;V1b). together with problem items *doesn't get along* (number 25) and *unhappy, sad or depressed* (number 103) represent extremely powerful predictors.

REFERENCES

1. Verhulst FC, Akkerhuis GW, Althaus M. Mental health in Dutch children: (I) a cross-cultural comparison. Acta Psychiatr Scand Suppl 1985;**323**:72.
2. Verhulst FC, Berden GF, Sanders-Woudstra JA. Mental health in Dutch children: (II). The prevalence of psychiatric disorder and relationship between measures. Acta Psychiatr Scand Suppl 1985;**324**:1–45.
3. Verhulst FC, Akkerhuis GW. Mental health in Dutch children: (III) behavioral-emotional problems reported by teachers of children aged 4–12. Acta Psychiatr Scand Suppl 1986;**330**:74.
4. American Psychiatric Association. American Psychiatric Association, editor. Diagnostic and statistical manual mental disorders. 4th ed. Washington, DC:1994.
5. World Health Organization. World Health Organization, editor. ICD-10: The ICD-10 Classification of Mental and Behavioral Disorders: Clinical Description and Diagnostic Guidelines. Geneva:1992.
6. Cantwell DP. Classification of child and adolescent psychopathology. J Child Psychol Psychiatry 1996;**37**:3–12.
7. Werry JS. Child psychiatric disorders: are they classifiable? Br J Psychiatry 1992;**161**:472–480.
8. Rutter M, Tuma AH, Lamm I. Assessment and Diagnosis in Child Psychopathology. London: David Fulton;1988.
9. Achenbach TM. Manual for the Child Behavior Checklist/4–18 and 1991 Profile. Burlington: University of Vermont; 1993.
10. Verhulst FC, et al. Empirically based assessment and taxonomy of psychopathology: cross-cultural applications. A review. Eur Child Adolesc Psychiatry 1995;**4**:61–76.
11. Brown JS, Achenbach TM. Bibliography of Published Studies Using the Child Behavior Checklist and Related Materials. 1994;Ed. Burlington, VT: University of Vermont Department of Psychiatry.
12. Achenbach TM. The classification of children's psychiatric symptoms: a factor-analytic study. Psychol Monogr;1966.
13. Achenbach TM. The Child Behavior Profile: I. Boys aged 6–11. J Consult Clin Psychol 1978;**46**:478–488.
14. Achenbach TM. Manual for the Teacher's Report Form and 1991 profile. Burlington: University of Vermont;1993.
15. Achenbach TM. Manual for the Youth Self-Report and 1991 Profile. Burlington, VT: University of Vermont Dept of Psychiatry;1993.
16. Stanger C, Fombonne E, Achenbach TM. Epidemiological comparisons of American and French children: Parent report of problems and competencies for ages 6–11. Eur Child Adolesc Psychiatry 1994;**3(1)**:16–28.
17. Macdonald VM, Tsiantis J, Achenbach TM, Motti-Stefanidi F, Richardson SC. Competencies and problems reported by parents of Greek and American children, ages 6–11. Eur Child Adolesc Psychiatry 1995;**4**:1–13.
18. Achenbach TM, Hensley VR, Phares V, Grayson D. Problems and competencies reported by parents of Australian and American children. J Child Psychol Psychiatry 1990;**31**:265–286.
19. Achenbach TM, Bird HR, Canino G, Phares V, Gould MS, Rubio-Stipec M. Epidemiological comparisons of Puerto Rican and U.S. mainland children: parent, teacher, and self-reports. J Am Acad Child Adolesc Psychiatry 1990;**29**:84–93.
20. Weine AM, Phillips JS, Achenbach TM. Behavioral and emotional problems among Chinese and American children: parent and teacher reports for ages 6 to 13. J Abnorm Child Psychol 1995;**23**:619–639.
21. Achenbach TM, et al. Epidemiological comparisons of American and Dutch children: I. Behavioral/emotional problems and competencies reported by parents for ages 4 to 16. J Am Acad Child Adolesc Psychiatry 1987;**26**:317–325.
22. Achenbach TM, Verhulst FC, Edelbrock C, Baron GD, Akkerhuis GW. Epidemiological comparisons of American and Dutch children: II. Behavioral/emotional problems reported by teachers for ages 6 to 11. J Am Acad Child Adolesc Psychiatry 1987;**26**:326–332.
23. Verhulst FC, Achenbach TM, Ferdinand RF, Kasius MC. Epidemiological comparisons of American and Dutch adolescents' self-reports. J Am Acad Child Adolesc Psychiatry 1993;**32**:1135–1144.
24. Achenbach TM, Verhulst FC, Baron GD, Althaus M. A comparison of syndromes derived from the Child Behavior Checklist for American and Dutch boys aged 6–11 and 12–16. J Child Psychol Psychiatry 1987;**28**:437–453.
25. Verhulst FC, Achenbach TM, Althaus M, Akkerhuis GW. A comparison of syndromes derived from the child behavior checklist for American and Dutch girls aged 6–11 and 12–16. J Child Psychol Psychiatry 1988;**29**:879–895.
26. Crijnen AA, Achenbach TM, Verhulst FC. Comparisons of problems reported by parents of children in 12 cultures: total problems, externalizing, and internalizing. J Am Acad Child Adolesc Psychiatry 1997;**36**:1269–1277.
27. Bird HR. Epidemiology of childhood disorders in a cross-cultural context. J Child Psychol Psychiatry 1996;**37**: 35–49.
28. De Groot A, Koot HM, Verhulst FC. Cross-cultural generalizability of the child behavior checklist cross-informant syndromes. Psychological Assessment 1994; **6(3)**:225–230.
29. Nunnally JC, Bernstein IH. Psychometric Theory. 3rd ed. New York: McGraw-Hill: 1994.
30. Streiner DL, Norman GR. Health Measurement Scales. A Practical Guide to their Development and Use. 2nd ed. Oxford: Oxford University Press;1995.
31. Larsson B, Frisk M. Social competence and emotional/behavioral problems in 6–16 year-old Swedish school children. Eur Child Adolesc Psychiatr 1999;**1**:24–33.
32. Steinhausen HC, Winkler Metzke C, Meier M, Kannenberg R. Behavioral and emotional problems reported by parents for ages 6 to 17 in a Swiss epidemiological study. Eur Child Adolesc Psychiatry 1997;**6**:136–141.
33. Hannesdottir H, Einarsdottir S. The Icelandic Child Mental Health Study. An epidemiological study of Icelandic children 2-18 years of age using the child behavior checklist as a screening instrument. Eur Child Adolesc Psychiatry 1995;**4**:237–248.
34. Nøvik TS. 1998; Validity of the Child Behavior Checklist in a Norwegian sample. Eur Child Adolesc Psychiatr (in press).
35. Sandberg DE, Meyer-Bahlburg HF, Yager TJ. The Child Behavior Checklist nonclinical standardization samples: should they be utilized as norms? J Am Acad Child Adolesc Psychiatry 1991;**30**:124–134.
36. Weisz JR, et al. Parent reports of behavioral and emotional problems among children in Kenya, Thailand, and the United States. Child Dev 1993;**64**:98–109.
37. Dopfner M, Pluck J, Berner W, Fegert JM, Huss M, Lenz

K, et al. [Mental disturbances in children and adolescents in Germany. Results of a representative study: age, gender and rater effects]. Z Kinder Jugendpsychiatr Psychother 1997;**25**:218–233. German.
38. Dopfner M, Schmeck K, Berner W, Lehmkuhl G, Poustka F. [Reliability and factorial validity of the Child Behavior Checklist — an analysis of a clinical and field sample (see comments)]. Z Kinder Jugendpsychiatr 1994;**22**:189–205. German.
39. Dopfner M, Berner W, Schmeck K, et al. Sergeant J, ed. Eunethydis. 1st ed. University of Amsterdam: Dept of Clinical Psychology, 1995. Internal consistency and validity of the CBCL and TRF in a German sample: A cross-cultural comparison. pp. 52–81.
40. Bird HR, Gould MS, Rubio-Stipec M, Staghezza BM, Canino G. Screening for childhood psychopathology in the community using the Child Behavior Checklist. J Am Acad Child Adolesc Psychiatry 1991;**30**:116–123.
41. Fombonne E. The Chartres Study: I. Prevalence of psychiatric disorders among French school-age children. Br J Psychiatry 1994;**164**:69–79.
42. Verhulst FC, Koot HM, van der Ende J. Differential predictive value of parents' and teachers' reports of children's problem behaviors: a longitudinal study. J Abnorm Child Psychol 1994;**22**:531–546.
43. Koot HM, Verhulst FC. Prediction of children's referral to mental health and special education services from earlier adjustment. J Child Psychol Psychiatry 1992;**33**:717–729.
44. Verhulst FC, et al. Pathways of problem behaviors from childhood to adolescence. J Am Acad Child Adolesc Psychiatry 1993;**32**:388–396.
45. Verhulst FC, van der Ende J. Six-year developmental course of internalizing and externalizing problem behaviors. J Am Acad Child Adolesc Psychiatry 1992;**31**:924–931.
46. Verhulst FC, van der Ende J. Six-year stability of parent-reported problem behavior in an epidemiological sample. J Abnorm Child Psychol 1992;**20**:595–610.
47. Verhulst FC, van der Ende J. 'Comorbidity' in an epidemiological sample: a longitudinal perspective. J Child Psychol Psychiatry 1993;**34**:767–783.
48. Verhulst FC, Koot HM, Berden GF. Four-year follow-up of an epidemiological sample. J Am Acad Child Adolesc Psychiatry 1990;**29**:440–448.
49. Verhulst FC, van der Ende J. Four-year follow-up of teacher-reported problem behaviors. Psychol Med 1991;**21**:965–977.
50. Boyle MH, Jones SC. Selecting measures of emotional and behavioral disorders of childhood for use in general populations. J Child Psychol Psychiatry 1985;**26**:137–159.
51. Fombonne E. The use of questionnaires in child psychiatry research: measuring their performance and choosing an optimal cut-off. J Child Psychol Psychiatry 1991; **32**:677–693.
52. Vaughan ML, Riccio CA, Hynd GW, Hall J. Diagnosing ADHD (predominantly inattentive and combined type subtypes): discriminant validity of the behavior assessment system for children and the Achenbach parent and teacher rating scales. J Clin Child Psychol 1997;**26**:349–357.
53. Chen WJ, Faraone SV, Biederman J, Tsuang MT. Diagnostic accuracy of the Child Behavior Checklist scales for attention-deficit hyperactivity disorder: a receiver-operating characteristic analysis. J Consult Clin Psychol 1994;**62**:1017–1025.
54. Biederman J, Faraone SV, Doyle A, Lehman BK, Kraus I, Perrin J, et al. Convergence of the Child Behavior Checklist with structured interview-based psychiatric diagnoses of ADHD children with and without comorbidity. J Child Psychol Psychiatry 1993;**34**:1241–1251.
55. Rey JM, Morris-Yates A. Adolescent depression and the Child Behavior Checklist. J Am Acad Child Adolesc Psychiatry 1991;**30**:423–427.
56. Stawski M, Auerbach JG, Barasch M, Lerner Y, Zimin R, Miller MS. Behavioral problems of adolescents with chronic physical illness: a comparison of parent-report and self-report measures. Eur Child Adolesc Psychiatry 1995;**4**:14–20.
57. Mehta M, Bagga A, Pande P, Bajaj G, Srivastava RN. Behavior problems in nephrotic syndrome. Indian Pediatr 1995;**32**:1281–1286.
58. van der Net J, Prakken AB, Helders PJ, ten Berge M, Van Herwaarden M, Sinnema G, et al. Correlates of disablement in polyarticular juvenile chronic arthritis — a cross-sectional study. Br J Rheumatol 1996;**35**:91–100.
59. Vandvik IH. Mental health and psychosocial functioning in children with recent onset of rheumatic disease. J Child Psychol Psychiatry 1990;**31**:961–971.
60. Engstrom I, Lindquist BL. Inflammatory bowel disease in children and adolescents: a somatic and psychiatric investigation. Acta Paediatr Scand 1991;**80**:640–647.
61. Utens EM, Verhulst FC, Meijboom FJ, Duivenvoorden HJ, Erdman RA, Bos E, et al. Behavioral and emotional problems in children and adolescents with congenital heart disease. Psychol Med 1993;**23**:415–424.
62. Curfs LM, Verhulst FC, Fryns JP. Behavioral and emotional problems in youngsters with Prader-Willi syndrome. Genet Couns 1991;**2**:33–41.
63. Van Dongen-Melman JE, et al. Commentary: potential pitfalls of using illness-specific measures. J Pediatr Psychol 1996;**21**:103–106.
64. Drotar D, Stein RE, Perrin EC. Methodological issues in using the Child Behavior Checklist and its related instruments in clinical child psychology research. J Clin Child Psychol 1995;**24(2)**:184–192.
65. Green A. [A demographic comparison of Funen and whole Denmark] En sammenlignende demografisk beskrivelse af Fyns amt og hele Danmark med henblik på vurdering af Fyns amts egnethed som repræsentativt udsnit af Danmark ved epidemiologiske studier. Odense, University of Odense, 1978.
66. Hansen EJ. Socio-economic classification in Denmark. Copenhagen: Danish Social Research Institute;1977.
67. Berg I, Fombonne E, McGuire R, Verhulst F. A cross cultural comparison of French and Dutch disturbed children using the Child Behavior Checklist (CBCL). Eur Child Adolesc Psychiatry 1997;**6**:7–11.
68. Duncan-Jones P, Grayson DA, Moran PAP. The utility of latent trait models in psychiatric epidemiology. Psychol Med 1986;**16**:391–405.
69. Rasch G. Pedagogic Institute in Denmark (eds). Studies in mathematical psychology I. Copenhagen 1960; Probalistic models for some intelligence and attainment tests.
70. Birnbaum A, Lord FM, Novick MR, (eds). Statistical theories of mental test scores. Some latent trait models and their use in inferring an examinee's ability. Addison-Wesley;1968; pp. 397–479.
71. Bilenberg N, Hørder K. [Behavioral symptoms among children and adolescents. Screening with the help of a questionnaire in a group of children aged 4 to 17 years]. Ugeskr Laeger 1998;**160**:4423–4428. Danish.
72. Fombonne E. Parent reports on behavior and competencies among 6–11-year-old French children. Eur Child Adolesc Psychiatry 1992;**1(4)**:233–243.
73. Lambert MC, et al. Epidemiology of behavioral and emotional problems among children of Jamaica and the United States: parent reports for ages 6 to 11. J Abnorm Child Psychol 1994;**22**:113–128.
74. Weisz JR, et al. Epidemiology of behavioral and emotional

problems among Thai and American children: parent reports for ages 6 to 11. J Am Acad Child Adolesc Psychiatry 1987;**26**:890–897.
75. Roussos A, et al. Achenbach's Child Behavior Checklist and Teachers' Report Form for Greek Children 6–12 years old. 1998; (pers.com.).
76. Poustka F, et al. [Psychological effects of military low altitude flight practice on children]. Z Kinder Jugendpsychiatr 1990;**18**:61 70. German.
77. Hellinckx W, Grietens H, Verhulst FC. Competence and behavioral problems in 6- to 12-year-old children in Flanders (Belgium) and Holland: A cross-national comparison. J Emotional and Behavioral Disorders 1994; **2(3)**:130–142.
78. Weisz JR, et al. Epidemiology of behavioral and emotional problems among Thai and American children: teacher reports for ages 6 11. J Child Psychol Psychiatry 1989;**30**:471–484.
79. Auerbach JG, Lerner Y. Syndromes derived from the Child Behavior Checklist for clinically referred Israeli boys aged 6–11: a research note. J Child Psychol Psychiatry 1991;**32**:1017–1024.
80. Verhulst FC, et al. Agreement between parents' and teachers' ratings of behavioral/emotional problems of children aged 4–12. J Child Psychol Psychiatry 1989; **30**:123–136.
81. Rey JM, Schrader E, Morris-Yates A. Parent-child agreement on children's behaviors reported by the Child Behavior Checklist (CBCL). J Adolesc 1992;**15**:219–230.
82. Verhulst FC, van der Ende J. Agreement between parents' reports and adolescents' self-reports of problem behavior. J Child Psychol Psychiatry 1992;**33**:1011–1023.

APPENDICES

Number of participants in every subgroup

Source	Sample	Boys 4–10 y	Boys 11–16 y	Girls 4–10 y	Girls 11–16 y
CBCL	population	N=217	N=141	N=226	N=196
	clinical	N=68	N=34	N=17	N=27
TRF	population	N=122	N=119	N=130	N=174
	clinical	N=53	N=35	N=11	N=19
YSR	population	N=0	N=140	N=0	N=196
	clinical	N=0	N=32	N=0	N=26

Supplementum

App. A1: School competence (parents and self rating)

Appendix A.1. Social competence scales CBCL & YSR

Item short text	Sample	Score (percentage distribution)				OR
V1: no. of friends		0	1	2–3	> =4	
CBCL n=781	population %	1	4	33	62	15.40*
n=146	clinical %	19	21	34	25	
YSR n=336	population %	0	2	18	80	3.88
n=58	clinical %	0	9	29	62	
V2: no. of contacts with friends per week			<1	1–2	> =3	
CBCL	population %		9	34	54	6.34*
	clinical %		32	27	37	0
YSR	population %		6	29	64	3.18
	clinical %		16	36	47	
VIa: behavior with sibs			worse	same	better	
CBCL	population %		4	55	32	5.42*
	clinical %		19	56	14	G
YSR	population %		7	49	35	1.34
	clinical %		12	59	19	G
VIb: behavior with others			worse	same	better	
CBCL	population %		1	46	52	
	clinical%		30	48	21	38.72*
YSR	population %		1	49	50	8.59
	clinical %		5	57	38	
VIc: behavior with parents			worse	same	better	
CBCL	population %		1	61	37	12.11*
	clinical %		12	62	23	
YSR	population %		2	49	48	7.69*
	clinical %		14	59	24	
VId: plays and works alone			worse	same	better	
CBCL	population %		5	46	48	4.22*
	clinical %		23	34	41	G
YSR	population %		7	50	42	1.36
	clinical %		10	50	36	

* Referred children do signif. worse than general population; O=age diff. signif., older socialize more than younger; G=sex diff. signif., girls behave better than boys (all $P<0.01$) bold line indicates 'cut off' in regression analysis.

App. A2: School competence (parents and self rating)

Appendix A.2. School competence scales CBCL & YSR

Item short text	Sample	Score (percentage distribution)				OR
VIIa: academic performance reading		failing	below average	average	above average	
CBCL n=570	population %	2	9	43	46	2.35*
n=122	clinical%	12	17	39	31	G,S
YSR n=333	population %	1	11	51	37	0.95
n=55	clinical %	6	11	54	29	G
VIIb: academic performance spelling		failing	below average	average	above average	
CBCL n=569	population %	3	14	47	36	2.23*
n=122	clinical %	15	21	41	23	G
YSR n=332	population %	3	19	43	35	1.10
n=55	clinical %	7	18	49	26	
VIIc: academic performance arithmetic, math.		failing	below average	average	above average	
CBCL n=573	population %	1	7	54	38	4.51*
n=124	clinical %	6	27	48	19	
YSR n=333	population %	2	14	48	36	0.93
n=55	clinical %	0	24	49	27	S
VIId: academic performance natural science		failing	below average	average	above average	
CBCL n=504	population %	1	7	67	25	2.31
n=102	clinical %	5	12	66	17	0
YSR n=299	population %	4	20	62	14	0.93
n=50	clinical %	12	14	58	16	
VIIe: academic performance English language		failing	below average	average	above average	
CBCL n=397	population %	2	9	55	34	1.31
n=74	clinical %	9	11	53	27	S
YSR n=331	population %	4	16	51	29	0.97
n=51	clinical %	4	22	45	29	S

* Referred children perform signif. worse than general population; O=age diff. signif., older perform better; G=sex diff. signif., girls perform better than boys; S=high SES perform signif. better (all $P<0.01$) bold line indicates 'cut off' in regression analysis.

Supplementum

App. A3: School competence (teacher rating)

Appendix A.3. School competence scales TRF

Item short text	Score (percentage distribution)							OR
VII: academic performance population n = 547 clinical n = 118		far below grade	somewhat below grade	at grade level	somewhat above grade	far above grade		
VII 1: reading								
population %		5	12	35	40	8		2.40*
clinical %		18	24	33	19	6		G,S
VII 2: spelling								
population %		7	14	39	36	4		3.75*
clinical %		29	28	27	14	2		G,S
VII 3: arithm., math.								
population %		1	13	46	36	3		4.01*
clinical %		10	35	35	20	0		S
VIII: compared to other pupils at the same age	much less	somewhat less	slightly less	above average	slightly more	somewhat more	much more	
VIII1: how hard is he or	2	5	12	36	22	20	3	4.58*
she working?	25	24	1	21	7	9	3	G,S
VIII2: how is he or she	1	3	7	40	21	23	5	9.03*
behaving?	18	23	19	19	6	12	3	G
VIII3: how much is he or	2	6	10	36	22	21	3	2.99*
she learning?	15	16	14	29	9	15	2	G
VIII4: how happy is he or she?	1	2	9	46	21	19	2	6.19*
	13	15	23	30	11	7	1	

* Referred children do signif. worse than general population; G = sex diff. signif., girls do better than boys; S = high SES do signif. better (all $P < 0.01$) bold line indicates 'cut off' in regression analysis.

Appendix B. Problem item scores and discriminative power (CBCL, TRF and YSR)

Item number and short text	Source	Sample	Item scores in % (+is frequency of item present=1+2)												Statistics odds ratios (OR) for single item discrimination adjusted for SES, age and sex *=significance at $P=0.01$	
			Boys						Girls							
			4–10 y			11–16 y			4–10 y			11–16 y				
			1	2	+	1	2	+	1	2	+	1	2	+	collapsed OR	stratified OR
N: *number of participants in ever subgroup*	CBCL	pop	N=217			N=141			N=226			N=196			abbreviations: B=boys score higher, G=girls score higher, Y=younger score higher, O=older score higher, S=low family SES score highe (all at signif. $P<0.01$) OR significance: *=significance $P<0.01$ For partial OR (¤=significance $P<0.05$)	
		clin	N=68			N=34			N=17			N=27				
	TRF	pop	N=122			N=119			N=130			N=174				
		clin	N=53			N=35			N=11			N=19				
	YSR	pop	N=0			N=140			N=0			N=196				
		clin	N=0			N=32			N=0			N=26				
1. acts too young	CBCL	pop	17	4	**21**	16	7	**23**	12	3	**15**	14	2	**16**	6.5*	yB 8.6* oB 5.3*
		clin	44	27	**71**	29	32	**61**	35	35	**70**	22	11	**33**	B	yG 14.6*
	TRF	pop	18	7	**25**	19	13	**32**	11	4	**15**	10	5	**15**	3.7*	
		clin	35	31	**66**	34	29	**63**	18	27	**45**	5	16	**21**	B	
	YSR	pop				29	4	**33**				27	1	**28**	1.3	
		clin				38	13	**51**				16	4	**20**		
2. allergy	CBCL	pop	5	7	**12**	14	9	**23**	8	8	**16**	13	5	**18**	1.2	
		clin	12	7	**19**	3	12	**15**	0	12	**12**	4	19	**23**		
	YSR	pop				8	14	**22**				13	7	**20**	0.9	
		clin				13	6	**19**				0	28	**28**		
3. argues a lot	CBCL	pop	52	18	**70**	53	26	**79**	52	22	**74**	52	23	**75**	1.1	
		clin	32	49	**81**	38	47	**85**	29	41	**70**	30	26	**56**		
	TRF	pop	26	15	**41**	33	19	**52**	25	8	**33**	34	16	**50**	1.4	
		clin	33	29	**62**	29	31	**60**	9	27	**36**	21	0	**21**		
	YSR	pop				49	29	**78**				54	17	**71**	0.7	
		clin				44	31	**75**				36	24	**60**		
4. asthma	CBCL	pop	4	5	**9**	6	6	**12**	4	3	**7**	3	3	**6**	1.6	
		clin	7	6	**13**	6	9	**15**	0	6	**6**	7	19	**26**		
	YSR	pop				4	6	**10**				3	4	**7**	1.9	
		clin				6	9	**15**				4	24	**28**		
5. acts like opposite sex	CBCL	pop	3	1	**4**	0	0	**0**	8	0	**8**	6	0	**6**	1.9	
		clin	6	3	**9**	3	0	**3**	0	6	**6**	7	0	**7**	G	
	TRF	pop	2	1	**3**	1	0	**1**	5	1	**6**	7	1	**8**	1.7	
		clin	4	4	**8**	0	0	**0**	9	0	**9**	0	0	**0**		
	YSR	pop				2	0	**2**				8	1	**9**	0.9	
		clin				6	3	**9**				4	4	**8**		
6. encopresis (BM outside toilet)	CBCL	pop	4	1	**5**	1	0	**1**	1	0	**1**	1	0	**1**	7.9*	yB 5.7* oB 16.2¤
		clin	9	15	**24**	12	3	**15**	0	6	**6**	0	0	**0**	B,Y	
7. bragging	CBCL	pop	29	7	**36**	37	2	**39**	26	2	**28**	10	2	**12**	2.1*	yB 2.4*
		clin	38	21	**59**	29	29	**58**	24	18	**42**	11	8	**19**	B,Y	
	TRF	pop	14	12	**26**	24	2	**26**	8	1	**9**	7	3	**10**	2.3*	
		clin	29	17	**46**	17	23	**40**	9	0	**9**	5	0	**5**	B,Y	
	YSR	pop				41	4	**45**				20	0	**20**	1.2	B
		clin				41	3	**44**				28	0	**28**		
8. can't concentrate	CBCL	pop	25	9	**34**	25	11	**36**	16	2	**18**	20	4	**24**	8.3*	yB 12.3* oB 7.8*
		clin	25	62	**87**	38	44	**82**	35	47	**82**	30	19	**49**	B,S	yG 23.5* oG 2.9¤
	TRF	pop	28	11	**39**	21	19	**40**	15	3	**18**	23	6	**29**	4.4*	
		clin	17	62	**79**	40	31	**71**	46	36	**82**	21	16	**37**	B,S	
	YSR	pop				34	14	**48**				41	6	**47**	1.2	
		clin				38	25	**63**				28	24	**52**	S	
9. can't get mind off thoughts (obsessions)	CBCL	pop	10	1	**11**	9	2	**11**	8	2	**10**	11	1	**12**	7.8*	yB 7.4* oB 8.2*
		clin	18	25	**43**	27	24	**51**	24	41	**65**	26	22	**48**		yG 14.7* oG 6.5*
	TRF	pop	0	2	**2**	3	0	**3**	4	0	**4**	2	1	**3**	10.4*	
		clin	21	14	**35**	20	9	**29**	9	0	**9**	21	0	**21**		
	YSR	pop				16	4	**20**				18	6	**24**	1.6	
		clin				22	6	**28**				4	36	**40**		

Item number and short text	Source	Sample	Item scores in % (+ is frequency of item present = 1 + 2)												Statistics odds ratios (OR) for single item discrimination adjusted for SES age and sex * = significance at $P = 0.01$	
			Boys						Girls							
			4–10 y			11–16 y			4–10 y			11–16 y				
			1	2	+	1	2	+	1	2	+	1	2	+	collapsed OR	stratified OR
10. can't sit still or hyperactive	CBCL	pop	30	5	**35**	24	4	**28**	12	2	**14**	9	2	**11**	5.9*	yB 5.6* oB 6.3*
		clin	29	46	**75**	41	32	**73**	24	41	**65**	26	15	**41**	B,S	yG 8.6* oG 4.3*
	TRF	pop	17	7	**24**	24	6	**30**	5	2	**7**	8	3	**11**	4.0*	
		clin	37	37	**74**	37	17	**54**	18	18	**36**	5	5	**10**	B	
	YSR	pop				41	11	**52**				31	8	**39**	1.1	
		clin				53	6	**59**				36	8	**44**	S	
11. too dependent	CBCL	pop	22	6	**28**	14	3	**17**	19	2	**21**	14	1	**15**	6.4*	yB 6.6* oB 5.6*
		clin	41	29	**70**	35	21	**56**	24	41	**65**	26	30	**56**	Y,S	yG 7.4* oG 6.2*
	TRF	pop	14	3	**17**	6	1	**7**	19	3	**22**	10	1	**11**	4.1*	
		clin	19	23	**42**	29	9	**38**	27	27	**54**	26	21	**47**	Y	
	YSR	pop				21	4	**25**				23	4	**27**	2.6*	
		clin				38	19	**57**				40	4	**44**		
12. lonely	CBCL	pop	14	2	**16**	9	1	**10**	13	0	**13**	12	1	**13**	5.5*	yB 4.0* oB 8.9*
		clin	25	21	**46**	32	21	**53**	35	18	**53**	19	22	**41**	S	yG 11.0* oG 5.6*
	TRF	pop	8	1	**9**	4	3	**7**	5	4	**9**	8	3	**11**	5.1*	
		clin	25	8	**33**	9	11	**20**	27	18	**45**	32	11	**43**		
	YSR	pop				16	4	**20**				27	3	**30**	2.0	
		clin				25	16	**41**				12	28	**40**		
13. confused	CBCL	pop	5	1	**6**	8	0	**8**	2	0	**2**	3	0	**3**	15.3*	yB 18.7* oB 9.8*
		clin	35	16	**51**	29	15	**44**	24	24	**48**	11	22	**33**	B,S	yG 46.1* oG 16.4*
	TRF	pop	8	2	**10**	9	3	**12**	4	1	**5**	6	2	**8**	3.6*	
		clin	23	21	**44**	17	6	**23**	18	9	**27**	16	5	**21**		
	YSR	pop				14	1	**15**				20	1	**21**	1.9	
		clin				25	3	**28**				12	20	**32**		
14. cries a lot	CBCL	pop	8	1	**9**	7	0	**7**	8	0	**8**	3	1	**4**	7.3*	yB 4.9* oB 6.4*
		clin	29	7	**36**	29	9	**38**	35	24	**59**	15	11	**26**	S	yG 21.2* oG 9.7*
	TRF	pop	5	2	**7**	5	0	**5**	5	1	**6**	5	1	**6**	3.4*	
		clin	14	8	**22**	14	0	**14**	9	18	**27**	11	0	**11**		
	YSR	pop				16	1	**17**				22	4	**26**	2.3	
		clin				28	9	**37**				32	20	**52**		
15. cruel to animals	CBCL	pop	1	0	**1**	1	0	**1**	0	0	**0**	1	0	**1**	17.2*	yB 18.1*
		clin	21	2	**23**	6	0	**6**	12	6	**18**	0	0	**0**		
16. mean to others	CBCL	pop	7	0	**7**	6	1	**7**	2	0	**2**	2	0	**2**	8.0*	yB 9.6* oB 3.8¤
		clin	32	12	**44**	18	6	**24**	29	6	**35**	7	0	**7**	B,S	yG 34.4*
	TRF	pop	11	0	**11**	8	1	**9**	3	0	**3**	5	2	**7**	5.2*	
		clin	23	15	**38**	26	9	**35**	0	18	**18**	26	0	**26**		
	YSR	pop				24	0	**24**				9	1	**10**	1.0	
		clin				13	0	**13**				28	0	**28**		
17. daydreams	CBCL	pop	23	7	**30**	32	4	**36**	26	3	**29**	36	3	**39**	3.5*	yB 2.8* oB 4.6*
		clin	31	27	**58**	47	24	**71**	53	29	**82**	48	15	**63**		yG 9.8*
	TRF	pop	23	3	**26**	33	3	**36**	24	4	**28**	24	5	**29**	4.4*	
		clin	29	35	**64**	31	20	**51**	27	36	**63**	63	5	**68**		
	YSR	pop				30	13	**43**				40	20	**60**	1.2	
		clin				38	19	**57**				40	16	**56**	G	
18. harms self	CBCL	pop	2	0	**2**	0	0	**0**	1	0	**1**	2	0	**2**	12.7*	yB 9.4*
		clin	12	0	**12**	9	0	**9**	18	0	**18**	19	7	**26**		yG 14.1* oG 13.9*
	TRF	pop	1	0	**1**	1	0	**10**	0	0	**0**	1	1	**2**	14.1*	
		clin	2	0	**2**	9	0	**9**	9	0	**9**	0	16	**16**		
	YSR	pop				1	0	**1**				3	2	**5**	5.4*	
		clin				3	3	**6**				8	20	**28**	G	
19. demands attention	CBCL	pop	38	6	**44**	25	4	**29**	32	3	**35**	22	3	**25**	8.6*	yB 6.6* oB 10.7*
		clin	22	63	**85**	44	35	**79**	53	41	**94**	37	26	**63**	Y	yG 24.3* oG 5.2*
	TRF	pop	17	14	**31**	16	8	**24**	12	5	**17**	18	5	**23**	5.2*	
		clin	23	58	**81**	17	40	**57**	18	46	**64**	42	11	**53**	B	
	YSR	pop				42	4	**46**				35	5	**40**	1.0	
		clin				41	13	**54**				8	24	**32**		

Item number and short text	Source	Sample	Item scores in % (+ is frequency of item present = 1 + 2)												Statistics odds ratios (OR) for single item discrimination adjusted for SES age and sex * = significance at $P = 0.01$	
			Boys						Girls							
			4–10 y			11–16 y			4–10 y			11–16 y				
			1	2	+	1	2	+	1	2	+	1	2	+	collapsed OR	stratified OR
20. destroys own things	CBCL	pop	12	2	**14**	6	3	**9**	4	0	**4**	1	1	**2**	8.7*	yB 5.6* oB 8.4*
		clin	28	24	**52**	32	15	**47**	24	18	**42**	15	4	**19**	B,Y	yG 18.7* oG 26.9*
	TRF	pop	3	2	**5**	4	0	**4**	1	1	**2**	0	1	**1**	10.9*	
		clin	17	15	**32**	17	14	**31**	9	9	**18**	5	0	**5**	B	
	YSR	pop				10	3	**13**				8	0	**8**	4.0*	
		clin				41	3	**44**				20	0	**20**		
21. destroys other things	CBCL	pop	4	0	**4**	4	1	**5**	2	0	**2**	2	0	**2**	10.4*	yB 20.3* oB 5.7*
		clin	27	15	**42**	21	9	**30**	6	6	**12**	7	0	**7**	B	
	TRF	pop	2	1	**3**	1	0	**1**	0	0	**0**	0	1	**1**	22.1*	
		clin	31	6	**37**	14	11	**25**	18	0	**18**	0	0	**0**		
	YSR	pop				4	0	**4**				2	0	**2**	1.8	
		clin				13	0	**13**				0	0	**0**		
22. disobeys at home	CBCL	pop	37	2	**39**	16	1	**17**	29	1	**30**	17	1	**18**	2.6*	yB 2.7* oB 6.5*
		clin	50	16	**66**	38	24	**62**	18	18	**36**	22	0	**22**	B,Y,S	
	YSR	pop				29	4	**33**				35	3	**38**	1.6	
		clin														
23. disobeys at school	CBCL	pop	13	1	**14**	14	2	**16**	2	0	**2**	8	1	**9**	5.1*	yB 6.5* oB 3.9*
		clin	41	12	**53**	32	15	**47**	24	12	**36**	11	0	**11**	B	yG 40.2*
	TRF	pop	7	4	**11**	11	2	**13**	2	0	**2**	6	2	**8**	6.3*	
		clin	27	25	**52**	29	20	**49**	18	9	**27**	5	0	**5**	B	
	YSR	pop				26	5	**31**				24	3	**27**	1.1	
		clin				31	16	**47**				20	0	**20**		
24. doesn't eat well	CBCL	pop	26	4	**30**	16	2	**18**	22	3	**25**	16	2	**18**	2.4*	yG 3.7¤ oG 3.3¤
		clin	18	24	**42**	24	6	**30**	29	24	**53**	19	15	**34**	Y	
	YSR	pop				32	9	**41**				36	12	**48**	1.4	
		clin				34	9	**43**				40	32	**72**		
25. doesn't get along	CBCL	pop	9	2	**11**	9	3	**12**	1	0	**1**	5	1	**6**	19.0*	yB 21.3* oB 12.3*
		clin	44	28	**72**	44	21	**65**	47	24	**71**	15	4	**19**	B	yG >100*
	TRF	pop	17	6	**23**	15	3	**18**	8	2	**10**	14	3	**17**	6.6*	
		clin	50	23	**73**	43	14	**57**	18	18	**36**	47	0	**47**	B	
	YSR	pop				7	1	**8**				7	1	**8**	4.7*	
		clin				28	6	**34**				24	8	**32**		
26. lacks guilt	CBCL	pop	15	3	**18**	11	6	**17**	14	3	**17**	13	4	**17**	5.3*	yB 7.2* oB 4.8*
		clin	44	19	**63**	29	24	**53**	35	35	**70**	22	4	**26**		yG 13.7*
	TRF	pop	11	4	**15**	13	5	**18**	8	1	**9**	10	3	**13**	3.7*	
		clin	19	33	**52**	11	31	**42**	18	18	**36**	16	5	**21**	B,S	
	YSR	pop				26	9	**35**				19	7	**26**	1.6	
		clin				34	19	**53**				20	12	**32**		
27. easily jealous	CBCL	pop	38	5	**43**	25	4	**29**	36	6	**42**	20	6	**26**	3.2*	yB 2.5* oB 2.8¤
		clin	21	46	**67**	29	24	**53**	41	35	**76**	41	19	**60**	Y,S	yG 5.5¤ oG 4.7*
	TRF	pop	10	7	**17**	5	2	**7**	12	1	**13**	10	4	**14**	4.2*	
		clin	15	23	**38**	17	17	**34**	9	18	**27**	21	0	**21**		
	YSR	pop				33	5	**38**				45	12	**57**	1.0	
		clin				22	16	**38**				32	24	**56**	G	
28. eats nonfood	CBCL	pop	2	1	**3**	2	1	**3**	2	0	**2**	4	1	**5**	2.0	
		clin	3	4	**7**	3	0	**3**	6	0	**6**	7	4	**11**		
	TRF	pop	0	0	**0**	0	0	**0**	1	0	**1**	2	1	**3**	7.2*	
		clin	2	2	**4**	3	0	**3**	0	0	**0**	16	0	**16**		
29. fears	CBCL	pop	30	7	**37**	17	5	**22**	33	7	**40**	19	6	**25**	1.9*	oB 2.7¤
		clin	21	21	**42**	27	18	**45**	29	24	**53**	19	22	**41**	Y	oG 2.5¤
	TRF	pop	6	2	**8**	2	1	**3**	6	0	**6**	3	2	**5**	6.1*	
		clin	8	12	**20**	6	14	**20**	18	18	**36**	16	5	**21**		
	YSR	pop				24	13	**37**				29	19	**48**	1.2	
		clin				25	9	**34**				12	40	**52**		
30. fears school	CBCL	pop	3	1	**4**	1	0	**1**	2	0	**2**	3	0	**3**	10.4*	yB 3.4¤ oB 45.5*
		clin	10	2	**12**	18	9	**27**	35	0	**35**	4	15	**19**		yG 31.3* oG 7.9*
	TRF	pop	2	0	**2**	3	0	**3**	2	0	**2**	1	1	**2**	3.5	
		clin	4	2	**6**	14	0	**14**	0	0	**0**	5	5	**10**		
	YSR	pop				4	0	**4**				4	0	**4**	5.2*	
		clin				13	6	**19**				12	4	**16**		

Supplementum

Item number and short text	Source	Sample	Boys 4–10 y 1	2	+	Boys 11–16 y 1	2	+	Girls 4–10 y 1	2	+	Girls 11–16 y 1	2	+	Statistics: odds ratios (OR) for single item discrimination adjusted for SES, age and sex * = significance at $P = 0.01$ collapsed OR	stratified OR
			Item scores in % (+ is frequency of item present = 1 + 2)													
31. fears impulses	CBCL	pop	25	5	**30**	38	1	**39**	28	2	**30**	36	4	**40**	2.0*	yB 2.5*
		clin	35	13	**48**	18	27	**45**	59	24	**83**	41	22	**63**	0	yG 10.1*
	TRF	pop	20	3	**23**	24	1	**25**	36	4	**40**	32	4	**36**	1.6	
		clin	19	6	**25**	11	26	**37**	46	18	**64**	42	11	**53**	G	
	YSR	pop				37	9	**46**				47	13	**60**	1.4	
		clin				34	16	**50**				32	40	**72**		
32. needs to be perfect	CBCL	pop	25	7	**32**	43	7	**50**	25	5	**30**	36	9	**45**	1.6	yB 2.0¤
		clin	24	24	**48**	27	27	**54**	18	47	**65**	37	30	**67**	0	yG 4.1¤
	TRF	pop	27	7	**34**	29	4	**33**	33	12	**45**	35	8	**43**	1.5	
		clin	21	12	**33**	14	26	**40**	18	27	**45**	42	32	**74**	G	
	YSR	pop				41	10	**51**				40	11	**51**	1.1	
		clin				38	16	**54**				20	36	**56**		
33. feels unloved	CBCL	pop	14	2	**16**	12	2	**14**	14	1	**15**	16	2	**18**	6.6*	yB 5.3* oB 13.0*
		clin	35	16	**51**	56	12	**68**	41	18	**59**	33	22	**55**	Y,S	yG 8.2* oG 5.9*
	TRF	pop	10	2	**12**	7	2	**9**	5	2	**7**	6	2	**8**	5.5*	
		clin	21	12	**33**	20	6	**26**	27	18	**45**	32	11	**43**		
	YSR	pop				12	4	**16**				21	3	**24**	2.4	
		clin				19	6	**25**				36	12	**48**	S	
34. feels persecuted	CBCL	pop	21	3	**24**	16	5	**21**	16	1	**17**	17	2	**19**	5.2*	yB 4.9* oB 8.9*
		clin	32	29	**61**	44	29	**73**	35	18	**53**	33	19	**52**	Y,S	yG 7.7* oG 3.7*
	TRF	pop	11	8	**19**	14	3	**17**	9	1	**10**	13	3	**16**	4.4*	
		clin	33	17	**50**	29	26	**55**	9	18	**27**	21	5	**26**	B,S	
	YSR	pop				22	4	**26**				19	1	**20**	1.3	
		clin				22	9	**31**				28	8	**36**		
35. feels worthless	CBCL	pop	11	1	**12**	16	2	**18**	6	0	**6**	15	2	**17**	8.9*	yB 7.6* oB 8.6*
		clin	27	25	**52**	50	18	**68**	47	24	**71**	41	22	**63**	0	yG 25.7* oG 7.5*
	TRF	pop	17	2	**19**	16	1	**17**	6	1	**7**	18	3	**21**	6.1*	
		clin	40	8	**48**	23	23	**46**	36	27	**63**	53	5	**58**		
	YSR	pop				10	3	**13**				17	2	**19**	1.8	
		clin				22	3	**25**				28	12	**40**		
36. accident-prone	CBCL	pop	7	1	**8**	10	2	**12**	9	0	**9**	6	1	**7**	4.4*	yB 4.6* oB 4.4*
		clin	24	9	**33**	35	6	**41**	12	12	**24**	22	4	**26**	S	oG 3.6¤
	TRF	pop	6	1	**7**	7	3	**10**	1	1	**2**	3	1	**4**	2.6*	
		clin	17	0	**17**	6	8	**15**	0	0	**0**	5	0	**5**	B	
	YSR	pop				23	5	**28**				17	5	**22**	1.1	
		clin				19	9	**28**				32	8	**40**		
37. fighting	CBCL	pop	13	1	**14**	8	1	**9**	2	0	**2**	4	1	**5**	7.9*	yB 7.4* oB 9.9*
		clin	40	18	**58**	32	18	**50**	29	12	**41**	4	0	**4**	B,S	yG 32.5*
	TRF	pop	20	7	**27**	9	3	**12**	2	0	**2**	2	3	**5**	4.7*	
		clin	29	27	**56**	31	17	**50**	18	18	**36**	0	0	**0**	B	
	YSR	pop				21	4	**25**				8	1	**9**	2.3	
		clin				28	19	**47**				16	0	**16**	B	
38. is teased	CBCL	pop	13	3	**16**	12	4	**16**	9	1	**10**	12	0	**12**	6.1*	yB 7.5* oB 7.3*
		clin	40	19	**59**	41	18	**59**	29	29	**58**	11	11	**22**	B,Y,S	yG 14.4*
	TRF	pop	11	3	**14**	11	2	**13**	6	0	**6**	7	2	**9**	3.5*	
		clin	31	8	**39**	20	17	**37**	18	9	**27**	11	0	**11**	B	
	YSR	pop				16	7	**23**				13	3	**16**	2.0	
		clin				25	16	**41**				20	12	**32**		
39. hangs around with children who get in trouble	CBCL	pop	15	2	**17**	17	4	**21**	11	0	**11**	20	2	**22**	3.2*	yB 3.8* oB 3.7*
		clin	31	13	**44**	35	18	**53**	18	18	**36**	19	19	**38**		yG 3.4¤
	TRF	pop	12	3	**15**	10	3	**13**	4	0	**4**	12	4	**16**	7.9*	
		clin	27	33	**60**	14	40	**54**	27	9	**36**	26	5	**31**		
	YSR	pop				24	6	**30**				31	7	**38**	3.1*	
		clin				38	28	**66**				40	20	**60**		
40. hear things that aren't there	CBCL	pop	1	2	**3**	1	1	**2**	4	0	**4**	2	0	**2**	5.9*	yB 5.2* oB 11.9*
		clin	7	3	**10**	15	3	**18**	0	0	**0**	4	11	**15**		oG 12.9*
	TRF	pop	2	1	**3**	0	0	**0**	0	0	**0**	0	1	**1**	2.0	
		clin	0	4	**4**	3	3	**6**	0	0	**0**	0	0	**0**		
	YSR	pop				8	1	**9**				8	1	**9**	3.3*	
		clin				22	0	**22**				12	20	**32**		

Item number and short text	Source	Sample	Item scores in % (+ is frequency of item present = 1 + 2)												Statistics odds ratios (OR) for single item discrimination adjusted for SES age and sex * = significance at P=0.01	
			Boys						Girls							
			4–10 y			11–16 y			4–10 y			11–16 y				
			1	2	+	1	2	+	1	2	+	1	2	+	collapsed OR	stratified OR
41. impulsive,	CBCL	pop	41	4	**45**	32	4	**36**	27	1	**28**	29	3	**32**	3.8*	yB 3.8* oB 9.0*
acts without		clin	31	47	**78**	38	44	**82**	6	47	**53**	37	7	**44**	B,S	yG 3.0¤
thinking	TRF	pop	30	6	**36**	27	5	**32**	13	1	**14**	18	4	**22**	4.0*	
		clin	40	42	**82**	29	31	**60**	9	36	**45**	21	11	**32**	B	
	YSR	pop				40	2	**42**				39	4	**43**	0.9	
		clin				41	9	**50**				28	8	**36**		
42. would rather	CBCL	pop	15	3	**18**	18	2	**20**	14	0	**14**	25	2	**27**	4.4*	yG 5.4* oB 4.9*
be alone		clin	38	12	**50**	44	12	**56**	47	12	**59**	30	11	**41**	O	yG 9.7*
	TRF	pop	15	5	**20**	17	4	**21**	12	2	**14**	13	2	**15**	3.6*	
		clin	29	14	**43**	34	11	**45**	46	9	**55**	32	0	**32**		
	YSR	pop				24	4	**28**				27	4	**31**	1.8	
		clin				31	16	**47**				32	4	**36**		
43. lying or	CBCL	pop	24	1	**25**	13	3	**16**	22	0	**22**	10	1	**11**	3.2*	yB 3.6* oB 3.9*
cheating		clin	38	16	**54**	24	21	**45**	35	12	**47**	15	7	**22**	Y,S	yG 3.3¤
	TRF	pop	7	3	**10**	7	2	**9**	8	0	**8**	3	2	**5**	6.1*	
		clin	25	14	**39**	31	20	**51**	27	0	**27**	16	5	**21**		
	YSR	pop				30	1	**31**				15	1	**16**	1.1	
		clin				16	13	**29**				20	4	**24**	B	
44. bites	CBCL	pop	15	7	**22**	15	14	**29**	15	8	**23**	21	12	**33**	1.5	
fingernails		clin	13	22	**35**	12	27	**39**	18	29	**47**	33	7	**40**	S	
	TRF	pop	4	2	**6**	3	1	**4**	3	2	**5**	5	3	**8**	4.6*	
		clin	6	10	**16**	23	6	**29**	0	9	**9**	11	0	**11**		
	YSR	pop				21	21	**42**				24	25	**49**	1.5	
		clin				16	44	**60**				24	28	**52**		
45. nervous	CBCL	pop	7	2	**9**	11	1	**12**	7	0	**7**	14	1	**15**	11.6*	yB 13.0* oB 14.6*
		clin	46	12	**58**	50	15	**65**	47	18	**65**	37	19	**56**	O,S	yG 20.3* oG 6.0*
	TRF	pop	15	5	**20**	15	4	**19**	9	1	**10**	18	3	**21**	7.0*	
		clin	37	15	**52**	43	23	**66**	64	9	**73**	53	16	**69**		
	YSR	pop				22	1	**23**				29	4	**33**	2.4*	
		clin				34	16	**50**				28	28	**56**		
46. nervous	CBCL	pop	4	1	**5**	6	3	**9**	0	0	**0**	2	1	**3**	10.2*	yB 10.6* oB 3.9*
movements		clin	21	15	**36**	24	9	**33**	12	24	**36**	7	15	**22**	B,S	yG > 100* oG 12.3*
	TRF	pop	3	2	**5**	4	4	**8**	3	0	**3**	0	2	**2**	5.8	
		clin	15	14	**29**	23	17	**40**	9	9	**18**	0	11	**11**	B	
	YSR	pop				6	1	**7**				6	3	**9**	1.9	
		clin				9	6	**15**				4	20	**24**		
47. nightmares	CBCL	pop	18	2	**20**	11	1	**12**	21	0	**21**	10	1	**11**	3.5*	yB 3.9* oB 3.1¤
		clin	29	19	**48**	24	6	**30**	29	6	**35**	19	7	**26**	Y	oG 3.5¤
	YSR	pop				17	1	**18**				31	2	**33**	1.9	
		clin				31	9	**40**				32	8	**40**		
48. not liked	CBCL	pop	6	1	**7**	8	2	**10**	1	0	**1**	3	2	**5**	14.1*	yB 15.3* oB 8.4*
		clin	43	10	**53**	41	9	**50**	35	24	**59**	15	4	**19**	B	yG > 100* oG 7.1*
	TRF	pop	11	7	**18**	13	3	**16**	9	1	**10**	15	4	**19**	4.7*	
		clin	40	15	**55**	37	14	**51**	9	27	**36**	26	5	**31**		
	YSR	pop				12	4	**16**				12	5	**17**	1.7	
		clin				28	0	**28**				24	4	**28**		
49. constipated	CBCL	pop	5	1	**6**	2	0	**2**	4	0	**4**	3	0	**3**	4.6*	oB 6.2¤
		clin	6	3	**9**	9	3	**12**	6	12	**18**	11	7	**18**		yG 5.5¤ oG 8.1*
50. too fearful	CBCL	pop	9	1	**10**	6	1	**7**	6	0	**6**	8	1	**9**	7.5*	yB 5.2* oB 5.9*
or anxious		clin	25	13	**38**	15	15	**30**	24	35	**59**	26	22	**48**		yG 19.7* oG 11.6*
	TRF	pop	12	2	**14**	8	0	**8**	11	2	**13**	9	1	**10**	5.4*	
		clin	19	12	**31**	31	17	**48**	36	9	**45**	37	5	**42**		
	YSR	pop				9	1	**10**				14	2	**16**	3.5*	
		clin				22	0	**22**				36	16	**52**		
51. dizzy	CBCL	pop	2	0	**2**	4	1	**5**	1	0	**1**	7	0	**7**	6.6*	yB 9.3* oB 5.3*
		clin	15	2	**17**	21	0	**21**	6	0	**6**	26	7	**33**	O	oG 6.7*
	TRF	pop	2	0	**2**	1	0	**1**	1	0	**1**	1	1	**2**	9.5*	
		clin	4	2	**6**	9	3	**12**	9	0	**9**	16	0	**16**		
	YSR	pop				13	2	**15**				22	1	**23**	2.5*	
		clin				28	0	**28**				44	8	**52**		

Item number and short text	Source	Sample	Item scores in % (+ is frequency of item present = 1 + 2)												Statistics odds ratios (OR) for single item discrimination adjusted for SES age and sex * = significance at $P = 0.01$	
			Boys						Girls							
			4–10 y			11–16 y			4–10 y			11–16 y				
			1	2	+	1	2	+	1	2	+	1	2	+	collapsed OR	stratified OR
52. feels too guilty	CBCL	pop	7	2	**9**	9	1	**10**	7	2	**9**	10	1	**11**	5.2*	yB 4.4* oB 5.2*
		clin	24	9	**33**	21	18	**39**	29	24	**53**	30	15	**45**		yG 13.7* oG 5.8*
	TRF	pop	6	0	**6**	1	1	**2**	4	0	**4**	4	2	**6**	3.0	
		clin	0	4	**4**	6	14	**20**	9	0	**9**	21	0	**21**		
	YSR	pop				15	6	**21**				25	5	**30**	1.4	
		clin				16	9	**25**				32	16	**48**		
53. eats too much	CBCL	pop	8	1	**9**	14	2	**16**	10	0	**10**	12	2	**14**	2.0*	oB 2.7¤
		clin	10	9	**19**	21	12	**33**	18	6	**24**	22	0	**22**	0,S	
	YSR	pop				20	9	**29**				20	13	**33**	1.2	
		clin				28	9	**37**				20	12	**32**		
54. overtired	CBCL	pop	13	1	**14**	11	1	**12**	14	0	**14**	16	1	**17**	3.4*	yB 2.8* oB 10.2*
		clin	24	7	**31**	38	15	**53**	24	6	**30**	15	22	**37**		
	TRF	pop	10	2	**12**	13	1	**14**	8	0	**8**	3	2	**5**	3.7*	
		clin	17	6	**23**	23	17	**40**	18	0	**18**	42	0	**42**		
	YSR	pop				27	5	**32**				32	7	**39**	2.5*	
		clin				41	9	**50**				52	16	**68**		
55. overweight	CBCL	pop	4	2	**6**	9	1	**10**	5	2	**7**	8	2	**10**	1.7	oB 3.8*
		clin	4	4	**8**	21	6	**27**	0	12	**12**	15	4	**19**	0,S	
	TRF	pop	1	4	**5**	6	1	**7**	6	2	**8**	8	5	**13**	2.5*	
		clin	6	8	**14**	11	9	**20**	9	18	**27**	5	0	**5**		
	YSR	pop				9	4	**13**				13	8	**21**	1.4	
		clin				19	6	**25**				8	8	**16**		
56a. aches, pains	CBCL	pop	7	1	**8**	4	1	**5**	5	1	**6**	6	1	**7**	3.4*	oB 10.2*
		clin	4	2	**6**	29	0	**29**	12	18	**30**	22	11	**33**		yG 9.1* oG 6.2*
	TRF	pop	5	1	**6**	3	1	**4**	3	1	**4**	2	1	**3**	1.0	
		clin	4	2	**6**	3	3	**6**	0	0	**0**	11	5	**16**		
	YSR	pop				9	1	**10**				10	1	**11**	3.7*	
		clin				28	6	**34**				16	12	**28**		
56b. headaches	CBCL	pop	13	1	**14**	19	3	**22**	12	1	**13**	22	2	**24**	2.4*	yB 2.2¤ oB 2.7¤
		clin	19	7	**26**	32	12	**44**	24	12	**36**	30	15	**45**	0	yG 4.2¤ oG 3.0¤
	TRF	pop	6	2	**8**	4	1	**5**	6	1	**7**	5	2	**7**	2.6*	
		clin	8	6	**14**	11	6	**17**	0	9	**9**	26	5	**31**		
	YSR	pop				21	4	**25**				26	7	**33**	2.1	
		clin				31	16	**47**				40	16	**56**		
56c. nausea, feels sick	CBCL	pop	4	1	**5**	3	0	**3**	2	0	**2**	7	0	**7**	5.2*	oB 13.1*
		clin	6	6	**12**	21	6	**27**	0	12	**12**	26	15	**41**		yG 11.1¤ oG 7.6*
	TRF	pop	2	1	**3**	0	1	**1**	3	0	**3**	3	1	**4**	3.5	
		clin	2	2	**4**	6	3	**9**	0	0	**0**	32	0	**32**		
	YSR	pop				14	1	**15**				10	0	**10**	4.2*	
		clin				31	6	**37**				32	20	**52**		
56d. eye problems	CBCL	pop	7	3	**10**	12	7	**19**	5	2	**7**	12	7	**19**	1.4	
		clin	6	6	**12**	15	15	**30**	6	6	**12**	22	0	**22**	0,S	
	TRF	pop	4	2	**6**	7	2	**9**	2	1	**3**	3	3	**6**	1.6	
		clin	6	2	**8**	17	6	**23**	0	0	**0**	5	0	**5**		
	YSR	pop				9	10	**19**				11	8	**19**	1.6	
		clin				16	13	**29**				12	24	**36**		
56e. skin problems	CBCL	pop	14	3	**17**	11	6	**17**	14	5	**19**	19	3	**22**	1.4	yB 2.2¤
		clin	16	10	**26**	12	6	**18**	18	18	**36**	11	4	**15**		
	TRF	pop	3	0	**3**	4	1	**5**	3	2	**5**	3	2	**5**	1.2	
		clin	4	4	**8**	3	0	**3**	0	9	**9**	0	11	**11**		
	YSR	pop				14	6	**20**				24	9	**33**	0.9	
		clin				13	9	**22**				20	12	**32**	G	
56f. stomach-aches	CBCL	pop	12	2	**14**	6	1	**7**	13	1	**14**	9	1	**10**	3.1*	oB 9.3*
		clin	18	2	**20**	32	9	**41**	18	12	**30**	22	15	**37**		oG 4.7*
	TRF	pop	3	2	**5**	3	1	**4**	5	1	**6**	3	1	**4**	1.5	
		clin	2	2	**4**	6	3	**9**	0	0	**0**	16	5	**21**		
	YSR	pop				9	0	**9**				16	2	**18**	3.2*	
		clin				22	6	**28**				28	12	**40**		

Item number and short text	Source	Sample	Boys 4–10 y 1	2	+	Boys 11–16 y 1	2	+	Girls 4–10 y 1	2	+	Girls 11–16 y 1	2	+	collapsed OR	stratified OR
			Item scores in % (+ is frequency of item present = 1 + 2)												Statistics odds ratios (OR) for single item discrimination adjusted for SES age and sex * = significance at P=0.01	
56g. vomiting	CBCL	pop	3	1	**4**	1	0	**1**	4	0	**4**	2	0	**2**	3.6*	oB 15.3¤
		clin	4	2	**6**	9	3	**12**	18	6	**24**	7	4	**11**		yG 9.5*
	TRF	pop	1	2	**3**	0	0	**0**	2	0	**2**	0	1	**1**	2.5	
		clin	2	0	**2**	6	0	**6**	0	0	**0**	0	5	**5**		
	YSR	pop				1	0	**1**				3	0	**3**	7.2*	
		clin				9	0	**9**				16	8	**24**		
57. attacks people	CBCL	pop	18	1	**19**	8	1	**9**	5	0	**5**	4	1	**5**	7.9*	yB 9.5* oB 4.7*
		clin	49	18	**67**	24	9	**33**	24	12	**36**	4	4	**8**	B,Y	yG 10.0*
	TRF	pop	20	6	**26**	11	1	**12**	2	0	**2**	6	2	**8**	5.4*	
		clin	39	29	**68**	26	23	**49**	27	9	**36**	0	0	**0**	B,S	
	YSR	pop				19	2	**21**				7	1	**8**	2.4	
		clin				34	9	**43**				12	4	**16**	B,S	
58. picking	CBCL	pop	26	7	**33**	16	4	**20**	29	4	**33**	13	3	**16**	2.6*	yB 2.2* oB 2.9¤
		clin	22	31	**53**	15	27	**42**	35	35	**70**	4	15	**19**	Y	yG 5.9*
	TRF	pop	2	2	**4**	4	1	**5**	2	1	**3**	1	1	**2**	4.9*	
		clin	14	10	**24**	9	6	**15**	0	9	**9**	5	0	**5**	B	
	YSR	pop				29	8	**37**				24	7	**31**	1.3	
		clin				28	9	**37**				16	20	**36**		
59. plays with private parts in public	CBCL	pop	7	0	**7**	0	1	**1**	1	1	**2**	1	0	**1**	2.5	oB 7.3¤
		clin	9	0	**9**	6	6	**12**	6	12	**18**	0	0	**0**	Y	
60. plays with private parts too much	CBCL	pop	4	1	**5**	2	1	**3**	1	0	**1**	1	0	**1**	3.6*	yB 3.1¤
		clin	9	3	**12**	9	3	**12**	12	12	**24**	0	0	**0**		yG 13.0*
61. poor school work	CBCL	pop	11	1	**12**	21	7	**28**	6	0	**6**	17	2	**19**	5.3*	yB 8.5* oB 4.7*
		clin	21	32	**53**	38	27	**65**	29	24	**53**	30	4	**34**	B,O	yG 20.1*
	TRF	pop	16	4	**20**	23	9	**32**	8	0	**8**	10	5	**15**	5.5*	
		clin	35	40	**75**	37	26	**63**	64	0	**64**	21	0	**21**	B	
	YSR	pop				32	7	**39**				26	4	**30**	1.6	
		clin				41	13	**54**				36	12	**48**		
62. clumsy	CBCL	pop	13	3	**16**	11	4	**15**	1	0	**1**	4	0	**4**	6.3*	yB 5.2* oB 6.1*
		clin	40	13	**53**	32	18	**50**	24	6	**30**	19	0	**19**	B	yG 51.0* oG 5.5¤
	TRF	pop	16	3	**19**	12	3	**15**	8	1	**9**	5	1	**6**	4.4*	
		clin	35	23	**58**	34	9	**43**	18	9	**27**	11	0	**11**		
	YSR	pop				12	4	**16**				12	2	**14**	1.1	
		clin				13	0	**13**				20	8	**28**		
63. prefers older kids	CBCL	pop	30	7	**37**	28	8	**36**	27	1	**28**	23	4	**27**	1.4	
		clin	34	18	**52**	35	9	**44**	18	6	**24**	26	19	**45**	B	
	TRF	pop	6	2	**8**	10	1	**11**	5	1	**6**	9	2	**11**	3.2*	
		clin	12	2	**14**	9	9	**18**	27	0	**27**	21	11	**32**		
	YSR	pop				47	9	**56**				43	12	**55**	0.9	
		clin				31	22	**53**				36	32	**68**		
64. prefers younger kids	CBCL	pop	19	2	**21**	16	6	**22**	21	0	**21**	15	2	**17**	3.4*	yB 4.5* oB 4.1*
		clin	34	21	**55**	38	21	**59**	18	29	**47**	19	15	**34**	S	yG 3.5¤
	TRF	pop	7	5	**12**	11	3	**14**	5	1	**6**	5	2	**7**	4.5	
		clin	21	14	**35**	14	17	**31**	0	27	**27**	11	16	**27**		
	YSR	pop				34	2	**36**				28	2	**30**	0.9	
		clin				31	6	**37**				24	12	**36**		
65. refuses to talk	CBCL	pop	4	1	**5**	3	1	**4**	2	0	**2**	2	0	**2**	4.7*	yB 3.9* oB 4.9¤
		clin	15	4	**19**	18	0	**18**	6	12	**18**	7	0	**7**		yG 9.9*
	TRF	pop	2	1	**3**	6	1	**7**	5	1	**6**	3	1	**4**	4.2*	
		clin	12	6	**18**	11	0	**11**	18	9	**27**	11	0	**11**		
	YSR	pop				8	1	**9**				4	1	**5**	2.2	
		clin				16	0	**16**				24	0	**24**		
66. repeats actions (compulsions)	CBCL	pop	4	0	**4**	1	0	**1**	2	0	**2**	1	0	**1**	13.0*	yB 8.5* oB 8.7¤
		clin	9	16	**25**	3	9	**12**	6	18	**24**	4	7	**11**	S	yG 15.2*
	TRF	pop	0	0	**0**	1	0	**1**	0	1	**1**	0	1	**1**	20.5*	
		clin	8	12	**20**	3	9	**12**	0	0	**0**	5	0	**5**		
	YSR	pop				4	0	**4**				3	1	**4**	6.2*	
		clin				9	13	**22**				8	16	**24**		

Item number and short text	Source	Sample	Item scores in % (+ is frequency of item present = 1 + 2)												Statistics odds ratios (OR) for single item discrimination adjusted for SES age and sex * = significance at $P=0.01$	
			Boys						Girls							
			4–10 y			11–16 y			4–10 y			11–16 y				
			1	2	+	1	2	+	1	2	+	1	2	+	collapsed OR	stratified OR
67. runs away from home	CBCL	pop	1	0	**1**	2	1	**3**	1	0	**1**	0	1	**1**	11.4*	yB 21.1* oB 4.6¤
		clin	15	3	**18**	3	9	**12**	0	6	**6**	11	4	**15**		oG 34.4*
	YSR	pop				2	0	**2**				2	0	**2**	11.0*	
		clin				13	3	**16**				8	4	**12**		
68. screams a lot	CBCL	pop	6	1	**7**	1	2	**3**	7	2	**9**	3	1	**4**	6.6*	yB 6.6* oB 7.0*
		clin	21	13	**34**	9	9	**18**	29	12	**41**	0	4	**4**	Y,S	yG 7.9*
	TRF	pop	5	2	**7**	1	0	**1**	2	0	**2**	2	1	**3**	7.6*	
		clin	23	12	**35**	14	6	**20**	9	18	**27**	0	0	**0**		
	YSR	pop				6	1	**7**				8	1	**9**	3.3*	
		clin				13	13	**26**				24	4	**28**		
69. secretive	CBCL	pop	15	1	**16**	26	4	**30**	16	1	**17**	29	3	**32**	2.5*	yB 3.9* oB 2.6¤
		clin	29	15	**44**	38	12	**50**	6	24	**30**	26	15	**41**	0	
	TRF	pop	11	2	**13**	8	3	**11**	12	4	**16**	12	3	**15**	3.9*	
		clin	19	6	**25**	23	17	**40**	18	18	**36**	21	21	**42**		
	YSR	pop				39	6	**45**				48	8	**56**	0.9	
		clin				19	16	**35**				48	20	**68**		
70. sees things that aren't there	CBCL	pop	2	0	**2**	0	0	**0**	4	0	**4**	1	0	**1**	4.3*	yB 4.3¤
		clin	7	2	**9**	3	3	**6**	6	0	**6**	4	0	**4**		
	TRF	pop	1	0	**1**	0	0	**0**	0	0	**0**	0	0	**0**	11.4	
		clin	2	8	**10**	0	3	**3**	0	0	**0**	0	0	**0**		
	YSR	pop				4	0	**4**				5	1	**6**	3.9*	
		clin				16	3	**19**				12	12	**24**		
71. self-conscious	CBCL	pop	26	3	**29**	31	1	**32**	37	2	**39**	37	2	**39**	1.1	
		clin	25	10	**35**	24	12	**36**	24	18	**42**	30	7	**37**	G,S	
	TRF	pop	24	3	**27**	14	3	**27**	25	4	**29**	21	3	**24**	1.0	
		clin	14	2	**16**	14	9	**23**	27	18	**45**	21	11	**32**		
	YSR	pop				35	7	**42**				46	13	**59**	0.6	
		clin				19	13	**32**				44	16	**60**	G	
72. sets fires	CBCL	pop	2	0	**2**	2	0	**2**	0	0	**0**	0	0	**0**	9.5*	yB 10.3* oB 8.4*
		clin	13	4	**17**	18	0	**18**	0	0	**0**	7	0	**7**	B	
	YSR	pop				2	1	**3**				1	0	**1**	2.3	
		clin				3	0	**3**				4	0	**4**		
73. sexual problems	CBCL	pop	1	0	**1**	0	0	**0**	1	0	**1**	0	0	**0**	26.2*	yG 38.1*
		clin	0	0	**0**	6	0	**6**	12	6	**18**	0	0	**0**		
74. showing off	CBCL	pop	40	7	**47**	32	6	**38**	38	6	**44**	15	2	**17**	1.9*	yB 2.3* oB 2.8¤
		clin	44	27	**71**	44	18	**62**	18	24	**42**	7	7	**14**		
	TRF	pop	25	5	**30**	29	5	**34**	5	1	**6**	9	0	**9**	2.7*	
		clin	27	35	**62**	37	14	**51**	9	0	**9**	5	0	**5**	B	
	YSR	pop				32	6	**38**				20	3	**23**	1.0	
		clin				28	16	**44**				20	0	**20**	B	
75. shy or timid	CBCL	pop	28	2	**30**	23	1	**24**	33	4	**37**	26	4	**30**	1.7*	yB 2.2¤
		clin	35	9	**44**	32	6	**38**	12	29	**41**	30	7	**37**		
	TRF	pop	22	2	**24**	17	3	**20**	33	5	**38**	21	6	**27**	1.5	
		clin	17	4	**21**	23	9	**32**	27	9	**36**	58	11	**69**	G	
	YSR	pop				26	2	**28**				32	10	**42**	1.3	
		clin				28	13	**41**				28	20	**48**	G	
76. sleeps little	CBCL	pop	10	2	**12**	13	1	**14**	6	1	**7**	8	0	**8**	3.8*	yB 5.2* oB 2.9¤
		clin	18	19	**37**	21	9	**30**	12	24	**36**	11	7	**18**	B	yG 5.5*
	YSR	pop				28	2	**30**				19	3	**22**	1.8	
		clin				28	19	**47**				20	16	**36**		
77. sleeps much	CBCL	pop	7	1	**8**	9	1	**10**	5	1	**6**	10	3	**13**	1.9	oB 3.3¤
		clin	9	7	**16**	12	15	**27**	6	6	**12**	0	15	**15**		
	YSR	pop				16	3	**19**				21	4	**25**	0.8	
		clin				13	6	**19**				16	4	**20**		
78. smears feces	CBCL	pop	1	0	**1**	0	0	**0**	0	0	**0**	0	0	**0**	18.2	
		clin	0	0	**0**	3	0	**3**	6	6	**12**	0	0	**0**		

Item number and short text	Source	Sample	Item scores in % (+ is frequency of item present = 1 + 2) Boys 4–10 y 1	2	+	Boys 11–16 y 1	2	+	Girls 4–10 y 1	2	+	Girls 11–16 y 1	2	+	Statistics odds ratios (OR) for single item discrimination adjusted for SES age and sex * = significance at $P=0.01$ collapsed OR	stratified OR
79. speech problem	CBCL	pop	6	5	**11**	5	1	**6**	7	1	**8**	2	0	**2**	3.1*	oB 3.5*
		clin	18	15	**33**	9	3	**12**	6	18	**24**	4	0	**4**	Y	yG 4.2¤
	TRF	pop	4	2	**6**	6	2	**8**	5	1	**6**	2	0	**2**	3.2*	
		clin	10	14	**24**	11	3	**14**	27	0	**27**	0	0	**0**		
	YSR	pop				11	2	**13**				8	1	**9**	1.0	
		clin				13	3	**16**				12	4	**16**		
80. stares blankly	CBCL	pop	1	0	**1**	4	1	**5**	3	1	**4**	4	1	**5**	12.0*	yB 37.2* oB 6.3*
		clin	28	9	**37**	24	3	**27**	35	6	**41**	19	4	**23**		yG 24.3* oG 4.4¤
	TRF	pop	7	2	**9**	8	3	**11**	10	1	**11**	6	0	**6**	8.1*	
		clin	21	19	**40**	34	0	**34**	27	18	**45**	47	0	**47**		
81. steals at home	CBCL	pop	2	0	**2**	3	0	**3**	0	0	**0**	3	0	**3**	6.6*	yB 4.7¤ oB 10.1*
		clin	7	0	**7**	18	3	**21**	24	0	**24**	0	0	**0**		
	YSR	pop				4	0	**4**				3	0	**3**	6.7*	
		clin				16	9	**25**				0	0	**0**		
82. steals outside home	CBCL	pop	1	0	**1**	0	0	**0**	0	0	**0**	1	0	**1**	21.6*	yB 10.2*
		clin	6	4	**10**	15	0	**15**	24	0	**24**	4	0	**4**		
	TRF	pop	2	0	**2**	2	0	**2**	1	0	**1**	0	0	**0**	8.5*	
		clin	6	4	**10**	6	6	**12**	0	0	**0**	0	0	**0**		
	YSR	pop				1	1	**2**				1	0	**1**	9.9*	
		clin				16	3	**19**				4	0	**4**		
83. stores up unneeded things	CBCL	pop	24	5	**29**	7	4	**11**	24	5	**29**	6	2	**8**	2.6*	
		clin	24	25	**49**	24	15	**39**	24	24	**48**	7	7	**14**	Y	yB 2.2* oB 4.4*
	TRF	pop	2	0	**2**	1	0	**1**	2	0	**2**	0	0	**0**	11.4*	
		clin	8	8	**16**	3	9	**12**	0	9	**9**	0	0	**0**		
	YSR	pop				19	9	**28**				17	10	**27**	1.6	
		clin				19	16	**35**				28	12	**40**		
84. strange behavior	CBCL	pop	3	2	**5**	4	0	**4**	3	0	**3**	2	0	**2**	15.9*	yB 11.6* oB 16.6*
		clin	21	18	**39**	21	18	**39**	29	24	**53**	7	11	**18**		yG 40.2* oG 17.3*
	TRF	pop	4	2	**6**	7	1	**8**	3	2	**5**	2	2	**4**	12.0*	
		clin	23	19	**42**	20	26	**46**	18	18	**36**	5	5	**10**		
	YSR	pop				14	5	**19**				16	4	**20**	2.0	
		clin				25	6	**31**				28	4	**32**		
85. strange ideas	CBCL	pop	4	1	**5**	4	1	**5**	1	0	**1**	2	0	**2**	12.8*	yB 8.5* oB 6.0*
		clin	14	13	**27**	12	12	**24**	29	12	**41**	11	11	**22**		yG 179.5* oG 25.4*
	TRF	pop	2	1	**3**	3	1	**4**	2	0	**2**	1	1	**2**	9.5*	
		clin	14	8	**22**	20	11	**31**	0	9	**9**	5	0	**5**		
strange thoughts	YSR	pop				8	4	**12**				12	4	**16**	3.1*	
		clin				9	16	**25**				16	20	**36**		
86. stubborn	CBCL	pop	30	5	**35**	32	2	**34**	25	2	**27**	30	2	**32**	4.4*	yB 4.2* oB 6.1*
		clin	41	27	**68**	56	21	**77**	41	41	**82**	30	22	**52**		yG 11.3*
	TRF	pop	15	5	**20**	19	6	**25**	18	2	**20**	17	5	**22**	3.5*	
		clin	27	19	**46**	34	29	**63**	36	18	**54**	42	0	**42**		
	YSR	pop				40	26	**66**				46	30	**76**	1.1	
		clin				34	38	**72**				28	48	**76**		
87. moody	CBCL	pop	30	2	**32**	40	3	**43**	34	2	**36**	49	5	**54**	6.6*	yB 8.2* oB 8.6*
		clin	40	40	**80**	50	38	**88**	35	47	**82**	44	26	**70**	O,S	yG 6.9*
	TRF	pop	22	4	**26**	19	7	**26**	18	2	**20**	28	8	**36**	6.2*	
		clin	39	37	**76**	46	34	**80**	9	36	**45**	47	16	**63**		
	YSR	pop				34	10	**44**				46	18	**64**	2.5*	
		clin				47	22	**69**				32	52	**84**	G	
88. sulks a lot	CBCL	pop	17	1	**18**	9	2	**11**	19	1	**20**	17	1	**18**	4.2*	yB 3.6* oB 7.1*
		clin	25	21	**46**	32	15	**47**	35	29	**64**	19	15	**34**		yG 9.1*
	TRF	pop	7	2	**9**	8	7	**15**	9	1	**10**	13	3	**16**	3.6*	
		clin	21	14	**35**	23	14	**37**	9	18	**27**	26	0	**26**		
89. suspicious	CBCL	pop	7	0	**7**	9	1	**10**	4	0	**4**	6	1	**7**	7.4*	yB 6.2* oB 5.0*
		clin	21	12	**33**	27	12	**39**	18	24	**42**	22	19	**41**	O,S	yG 15.0* oG 14.2*
	TRF	pop	7	2	**9**	10	2	**12**	4	1	**5**	5	1	**6**	3.4*	
		clin	29	2	**31**	11	11	**22**	9	9	**18**	5	5	**10**		
	YSR	pop				34	6	**40**				33	4	**37**	1.7	
		clin				41	19	**60**				40	12	**52**		

Supplementum

Item number and short text	Source	Sample	Item scores in % (+ is frequency of item present = 1 + 2) Boys 4–10 y 1	2	+	Boys 11–16 y 1	2	+	Girls 4–10 y 1	2	+	Girls 11–16 y 1	2	+	Statistics odds ratios (OR) for single item discrimination adjusted for SES, age and sex * = significance at $P=0.01$ collapsed OR	stratified OR
90. swearing	CBCL	pop	43	2	**45**	33	4	**37**	29	1	**30**	26	2	**28**	1.8*	yG 6.6*
		clin	37	22	**59**	35	21	**56**	53	18	**71**	15	15	**30**	B,S	
	TRF	pop	10	4	**14**	7	1	**8**	5	0	**5**	11	2	**13**	4.3*	
		clin	27	17	**44**	20	29	**49**	0	18	**18**	5	0	**5**		
	YSR	pop				51	12	**63**				54	8	**62**	1.1	
		clin				47	25	**72**				40	24	**64**		
91. suicidal talk	CBCL	pop	3	1	**4**	1	0	**1**	1	1	**2**	1	1	**2**	16.3*	yB 14.9* oB 39.5*
		clin	24	10	**34**	29	3	**32**	29	12	**41**	11	4	**15**		yG 30.9* oG 9.1*
	TRF	pop	1	0	**1**	0	0	**0**	0	0	**0**	1	1	**2**	40.6*	
		clin	8	0	**8**	6	6	**12**	0	9	**9**	11	5	**16**		
suicidal thoughts	YSR	pop				5	1	**6**				7	1	**8**	4.4*	
		clin				13	6	**19**				24	12	**36**		
92. talks or walks in sleep	CBCL	pop	14	1	**15**	11	4	**15**	15	3	**18**	10	4	**14**	1.8*	yB 2.1¤ oB 2.8¤
		clin	13	12	**25**	21	15	**36**	29	0	**29**	15	7	**22**		
93. talks too much	CBCL	pop	31	9	**40**	30	4	**34**	29	4	**33**	23	4	**27**	2.0*	yG 9.7*
		clin	22	37	**59**	38	15	**53**	47	29	**76**	7	22	**29**	S	
	TRF	pop	19	10	**29**	23	5	**28**	13	2	**15**	12	3	**15**	2.1*	
		clin	31	25	**56**	23	17	**40**	9	27	**36**	11	0	**11**	B	
	YSR	pop				45	21	**66**				41	23	**64**	0.9	
		clin				50	16	**66**				40	24	**64**		
94. teases a lot	CBCL	pop	37	5	**42**	19	5	**24**	23	1	**24**	17	1	**18**	3.2*	yB 3.5* oB 6.3*
		clin	37	37	**74**	44	21	**65**	24	18	**42**	19	7	**26**	B,Y,S	
	TRF	pop	14	10	**24**	17	3	**20**	5	2	**7**	9	1	**10**	4.6*	
		clin	35	29	**64**	31	26	**57**	18	9	**27**	11	0	**11**	B	
	YSR	pop				31	6	**37**				22	2	**24**	0.8	
		clin				34	3	**37**				28	0	**28**	B	
95. temper tantrums	CBCL	pop	30	5	**35**	16	4	**20**	21	1	**22**	17	1	**18**	5.6*	yB 4.4* oB 10.7*
		clin	35	37	**72**	32	41	**73**	35	24	**59**	41	7	**48**	B,Y	yG 5.5* oG 4.4*
	TRF	pop	9	7	**16**	12	6	**18**	6	0	**6**	5	2	**7**	4.2*	
		clin	21	33	**54**	26	29	**55**	0	18	**18**	5	0	**5**	B,S	
	YSR	pop				31	13	**44**				35	11	**46**	1.9	
		clin				34	41	**75**				20	36	**56**		
96. thinks about sex too much	CBCL	pop	6	1	**7**	8	0	**8**	4	0	**4**	4	1	**5**	1.7	oB 3.6¤
		clin	6	3	**9**	18	6	**24**	12	6	**18**	4	0	**4**		
	TRF	pop	6	0	**6**	3	2	**5**	1	0	**1**	11	1	**12**	4.3*	
		clin	15	12	**27**	11	11	**22**	18	0	**18**	1 1	5	**16**		
	YSR	pop				21	6	**27**				15	2	**17**	0.7	
		clin				19	3	**22**				16	4	**20**		
97. threatens people	CBCL	pop	3	0	**3**	4	0	**4**	1	0	**1**	1	0	**1**	6.5*	yB 8.9* oB 5.7*
		clin	12	7	**19**	21	3	**24**	12	6	**18**	4	0	**4**		yG 12.6*
	TRF	pop	5	1	**6**	3	0	**3**	0	0	**0**	1	1	**2**	9.1*	
		clin	15	12	**27**	17	20	**37**	9	9	**18**	0	5	**5**	B	
	YSR	pop				9	2	**11**				2	1	**3**	2.7	
		clin				19	9	**28**				8	0	**8**		
98. thumbsucking	CBCL	pop	2	1	**3**	1	1	**2**	5	1	**6**	1	0	**1**	3.7*	yB 5.8*
		clin	13	7	**20**	3	3	**6**	12	0	**12**	4	0	**4**	Y	
99. too concerned with neat or clean	CBCL	pop	5	1	**6**	9	0	**9**	5	1	**6**	12	2	**14**	2.0	yG 4.5¤
		clin	9	3	**12**	6	3	**9**	18	6	**24**	11	11	**22**	0	yG 4.5¤
	TRF	pop	2	0	**2**	1	0	**1**	6	2	**8**	5	1	**6**	6.1*	
		clin	6	2	**8**	3	9	**12**	18	9	**27**	26	0	**26**	G	
	YSR	pop				26	3	**29**				31	6	**37**	1.0	
		clin				28	3	**31**				24	8	**32**		
100. trouble sleeping	CBCL	pop	4	1	**5**	6	0	**6**	7	0	**7**	4	2	**6**	10.2*	yB 14.3* oB 10.9*
		clin	22	16	**38**	21	18	**39**	35	12	**47**	7	15	**22**		yG 11.3* oG 9.9*
	YSR	pop				11	1	**12**				12	4	**16**	3.5*	
		clin				28	3	**31**				20	24	**44**		

Item number and short text	Source	Sample	Item scores in % (+ is frequency of item present = 1 + 2)												Statistics odds ratios (OR) for single item discrimination adjusted for SES, age and sex * = significance at $P = 0.01$	
			Boys						Girls							
			4–10 y			11–16 y			4–10 y			11–16 y				
			1	2	+	1	2	+	1	2	+	1	2	+	collapsed OR	stratified OR
101. truancy	CBCL	pop	1	0	**1**	4	0	**4**	1	0	**1**	4	1	**5**	3.4*	oB 7.5*
		clin	2	0	**2**	15	6	**21**	12	0	**12**	0	7	**7**	0	
	TRF	pop	3	0	**3**	1	0	**1**	0	0	**0**	1	1	**2**	7.9*	
		clin	10	0	**10**	6	3	**9**	0	0	**0**	11	5	**16**		
	YSR	pop				6	0	**6**				8	1	**9**	1.7	
		clin				3	3	**6**				16	0	**16**		
102. lacks energy	CBCL	pop	8	1	**9**	12	1	**13**	3	0	**3**	13	2	**15**	2.6*	yB 2.3¤ oB 4.4*
		clin	15	4	**19**	24	12	**36**	6	6	**12**	11	11	**22**	0	
	TRF	pop	8	3	**11**	11	3	**14**	12	3	**15**	8	2	**10**	2.1*	
		clin	2	14	**16**	14	11	**25**	9	0	**9**	42	0	**42**		
	YSR	pop				22	3	**25**				28	5	**33**	1.5	
		clin				34	0	**34**				28	24	**52**		
103. unhappy, sad, depressed	CBCL	pop	4	1	**5**	9	1	**10**	6	0	**6**	7	1	**8**	16.1*	yB 17.7* oB 16.5*
		clin	37	12	**49**	53	15	**68**	53	12	**65**	33	22	**55**	0	yG 29.0* oG 17.1*
	TRF	pop	11	1	**12**	8	3	**11**	5	1	**6**	10	1	**11**	11.0*	
		clin	37	10	**47**	37	17	**54**	27	9	**36**	58	21	**79**		
	YSR	pop				16	0	**16**				27	1	**28**	3.5*	
		clin				44	6	**50**				36	16	**52**		
104. unusually loud	CBCL	pop	16	2	**18**	4	2	**6**	10	0	**10**	6	1	**7**	5.0*	yB 5.7* oB 7.0*
		clin	31	27	**58**	24	12	**36**	18	12	**30**	15	4	**19**	B,Y,S	yG 3.5¤
	TRF	pop	7	5	**12**	13	3	**16**	2	0	**2**	8	2	**10**	5.4*	
		clin	37	17	**54**	14	26	**40**	9	18	**27**	11	0	**11**	B	
	YSR	pop				19	1	**20**				17	2	**19**	1.9	
		clin				22	16	**38**				32	4	**36**		
105. alcohol, drugs	CBCL	pop	1	0	**1**	3	0	**3**	0	0	**0**	11	1	**12**	1.2	
		clin	0	0	**0**	12	0	**12**	0	0	**0**	11	0	**11**		
	TRF	pop	0	0	**0**	1	0	**1**	0	0	**0**	0	0	**0**	>100	
		clin	0	0	**0**	6	3	**9**	0	0	**0**	0	11	**11**		
	YSR	pop				15	3	**18**				22	5	**27**	1.0	
		clin				19	3	**22**				28	0	**28**		
106. vandalism	CBCL	pop	2	0	**2**	3	1	**4**	0	0	**0**	0	0	**0**	12.8*	yB 19.0* oB 4.4¤
		clin	25	7	**32**	12	6	**18**	0	6	**6**	4	4	**8**		
107. wets during day	CBCL	pop	5	0	**5**	0	0	**0**	4	1	**5**	0	0	**0**	5.5*	
		clin	7	4	**11**	3	0	**3**	12	0	**12**	4	0	**4**	Y	
108. wets bed	CBCL	pop	10	5	**15**	4	2	**6**	6	2	**8**	1	1	**2**	2.6*	yG 7.3*
		clin	6	15	**21**	3	3	**6**	18	18	**36**	0	7	**7**	Y	
109. whining	CBCL	pop	5	1	**6**	5	0	**5**	7	0	**7**	3	0	**3**	4.3*	yB 4.6*
		clin	15	4	**19**	15	0	**15**	29	6	**35**	4	4	**8**		yG 8.6*
	TRF	pop	2	1	**3**	4	1	**5**	5	1	**6**	4	1	**5**	3.1*	
		clin	14	4	**18**	9	3	**12**	9	9	**18**	0	0	**0**		
110. wishes to be opposite sex	CBCL	pop	1	0	**1**	1	0	**1**	1	0	**1**	2	1	**3**	1.3	
		clin	0	2	**2**	0	0	**0**	0	0	**0**	4	0	**4**		
	YSR	pop				3	0	**3**				10	2	**12**	2.1	
		clin				3	0	**3**				20	8	**28**		
111. withdrawn	CBCL	pop	5	0	**5**	11	1	**12**	2	0	**2**	8	0	**8**	6.0*	yB 7.9* oB 2.7¤
		clin	19	9	**28**	18	12	**30**	18	12	**30**	30	7	**37**	0	yG 28.3* oG 6.4*
	TRF	pop	4	3	**7**	8	3	**11**	8	0	**8**	9	1	**10**	5.7*	
		clin	17	12	**29**	11	14	**25**	55	9	**64**	53	11	**64**		
	YSR	pop				35	11	**46**				42	7	**49**	1.9	
		clin				56	22	**78**				36	12	**48**		
112. worrying	CBCL	pop	18	1	**19**	23	1	**24**	18	3	**21**	29	3	**32**	4.4*	yB 4.7* oB 5.9*
		clin	32	21	**53**	47	18	**65**	29	35	**64**	37	19	**56**	0	yG 6.9* oG 2.7¤
	TRF	pop	19	5	**24**	12	2	**14**	22	2	**24**	20	2	**22**	3.9*	
		clin	27	10	**37**	31	11	**42**	55	9	**64**	47	16	**63**		
	YSR	pop				19	2	**21**				30	7	**37**	4.0*	
		clin				34	16	**50**				40	44	**84**		

Supplementum

Problem item scores and discriminative power (special TRF items)

Item number and short text	Source	Sample	Boys 4–10 y 1	Boys 4–10 y 2	Boys 4–10 y +	Boys 11–16 y 1	Boys 11–16 y 2	Boys 11–16 y +	Girls 4–10 y 1	Girls 4–10 y 2	Girls 4–10 y +	Girls 11–16 y 1	Girls 11–16 y 2	Girls 11–16 y +	Statistics: odds ratios (OR) for single item discrimination adjusted for SES, age and sex; * = significance at $P=0.01$; OR in collapsed group
N: number of participants in every subgroup	TRF	pop		N = 122			N = 119			N = 130			N = 174		
		clin		N = 53			N = 35			N = 11			N = 19		
2: hums, makes	TRF	pop	8	2	**10**	6	3	**9**	0	1	**1**	2	1	**3**	6.3* Bc
noises		clin	29	19	**48**	17	14	**31**	9	9	**18**	5	0	**5**	
4: fails to finish	TRF	pop	24	11	**35**	28	10	**38**	17	5	**22**	11	5	**16**	4.6* B
things		clin	21	52	**73**	43	26	**69**	55	18	**73**	16	5	**21**	
6: defiant, talks	TRF	pop	15	4	**19**	19	5	**24**	12	2	**14**	19	7	**26**	3.3*
back to staff		clin	37	23	**60**	20	26	**46**	18	18	**36**	26	0	**26**	
15: fidgets	TRF	pop	10	2	**12**	9	3	**12**	4	0	**4**	5	1	**6**	5.2* B
		clin	27	25	**52**	29	17	**46**	18	18	**36**	5	5	**10**	
22: difficulty	TRF	pop	16	9	**25**	19	5	**24**	11	2	**13**	15	2	**17**	4.8* B
following directions		clin	23	50	**73**	17	29	**46**	27	55	**82**	16	5	**21**	
24: disturbs other	TRF	pop	26	8	**34**	27	8	**35**	11	2	**13**	19	3	**22**	2.9* B
pupils		clin	25	50	**75**	31	23	**54**	36	9	**45**	0	5	**5**	
47: overconforms	TRF	pop	19	4	**23**	23	3	**26**	30	3	**33**	22	5	**27**	2.1* G
to rules		clin	19	15	**34**	11	11	**22**	36	18	**54**	53	11	**64**	
49: difficulty	TRF	pop	11	10	**21**	15	13	**28**	8	2	**10**	14	8	**22**	4.2* B
learning		clin	25	40	**65**	20	34	**54**	27	36	**63**	16	5	**21**	
53: talks out of	TRF	pop	21	12	**33**	24	9	**33**	13	4	**17**	14	6	**20**	2.4* B
turn		clin	35	37	**72**	11	34	**45**	27	18	**45**	0	0	**0**	
59: sleeps in class	TRF	pop	3	1	**4**	13	1	**14**	3	0	**3**	3	1	**4**	1.5 B,O
		clin	4	6	**10**	14	0	**14**	0	0	**0**	5	0	**5**	
60: apathetic or	TRF	pop	12	0	**12**	17	4	**21**	7	0	**7**	11	2	**13**	3.6*
unmotivated		clin	21	21	**42**	34	6	**40**	27	0	**27**	37	0	**37**	
67: disrupts class	TRF	pop	10	9	**19**	19	5	**24**	6	1	**7**	10	2	**12**	4.6* B
discipline		clin	19	50	**69**	34	20	**54**	9	18	**27**	5	5	**10**	
72: messy work	TRF	pop	20	7	**27**	32	9	**41**	12	0	**12**	13	3	**16**	3.5* B
		clin	31	35	**66**	26	29	**55**	46	9	**55**	26	5	**31**	
73: behaves	TRF	pop	11	3	**14**	3	3	**6**	1	0	**1**	3	1	**4**	7.3* B
irresponsibly		clin	23	25	**48**	23	17	**40**	18	9	**27**	5	0	**5**	
76: explosive and	TRF	pop	9	2	**11**	11	3	**14**	3	0	**3**	4	1	**5**	5.9* B
unpredictable behavior		clin	21	33	**54**	20	31	**51**	9	18	**27**	5	0	**5**	
77: demands must	TRF	pop	14	5	**19**	10	4	**14**	7	2	**9**	10	2	**12**	5.1* B
be met immediately		clin	27	37	**64**	23	26	**49**	0	27	**27**	21	0	**21**	
78: inattentive,	TRF	pop	24	10	**34**	26	9	**35**	12	2	**14**	16	4	**20**	4.3* B,S
easily distracted		clin	21	60	**81**	26	34	**60**	36	27	**63**	21	11	**32**	
81: feels hurt	TRF	pop	28	7	**35**	21	8	**29**	30	6	**36**	28	5	**33**	3.8*
when criticized		clin	29	33	**62**	40	26	**66**	46	18	**64**	47	16	**63**	
92: underachiev-	TRF	pop	11	6	**17**	24	4	**28**	7	0	**7**	10	5	**15**	6.6* B
ing, not working up to potential		clin	35	29	**64**	29	34	**63**	36	9	**45**	26	11	**37**	
98: tardy to school	TRF	pop	7	1	**8**	8	0	**8**	5	1	**6**	6	2	**8**	4.4*
or class		clin	14	12	**26**	9	11	**20**	36	0	**36**	5	11	**16**	
100: fails to carry	TRF	pop	13	3	**16**	17	7	**24**	7	0	**7**	14	2	**16**	3.9* B,S
out assigned tasks		clin	21	34	**55**	23	17	**40**	46	9	**55**	5	11	**16**	
106: overly anxious	TRF	pop	15	7	**22**	12	3	**15**	13	3	**16**	13	1	**14**	2.6*
to please		clin	21	15	**36**	26	11	**37**	27	18	**45**	16	21	**37**	
107: dislikes	TRF	pop	8	0	**8**	8	3	**11**	2	0	**2**	10	1	**11**	4.1* O
school		clin	21	8	**29**	26	9	**35**	9	0	**9**	16	5	**21**	
108: is afraid of	TRF	pop	19	7	**26**	14	2	**16**	28	5	**33**	24	3	**27**	2.5* G
making mistakes		clin	21	8	**29**	20	26	**46**	27	18	**45**	42	16	**58**	
110: unclean	TRF	pop	3	1	**4**	2	3	**5**	2	0	**2**	2	1	**3**	3.3* S
personal appearance		clin	8	4	**12**	9	0	**9**	9	0	**9**	16	0	**16**	