

Supplementum No. 330. Vol. 74. 1986

Mental health in
Dutch children:
(III) behavioral-emotional
problems reported by teachers
of children aged 4–12

Frank C. Verhulst
Grard W. Akkerhuis

MUNKSGAARD COPENHAGEN



ISBN 87-16-06369-4
ISSN 0065-1508

From the department of child psychiatry, Academic Hospital Rotterdam – Sophia Children's Hospital, Gordelweg 160, 3038 GE Rotterdam, The Netherlands

This research was made possible by financial support from the Sophia Foundation for Medical Research

INTRODUCTION

Since Lapouse and Monk (1) published the first true child psychiatric epidemiological study, research and practice in the field of child psychopathology have benefitted from the study of representative samples of normal and disturbed children in a number of ways. Several studies have provided base-line data on the prevalence of children's behavioral-emotional problems in the general population (for an overview, see Verhulst, Akkerhuis and Althaus) (2), whereas others have made detailed comparisons between normal and clinically referred children (3). Because these studies indicate that child psychiatric problems vary with demographic variables such as age, gender and socio-economic status, assessment of child psychopathology needs to take account of demographic differences.

Although population-based surveys are designed to determine the prevalence of children's behavioral - emotional problems, they also have implications for other issues in the study of child psychopathology. These issues are:

- (1) The quantitative nature of most childhood behavioral-emotional problems;
- (2) The need for a common set of standardized assessment procedures to facilitate comparison of findings across different studies carried out in different locations;
- (3) The dependence on adult informants for data needed to evaluate a child's condition;
- (4) The roles of situational specificity and informant variance in the assessment of behavioral-emotional problems.

The quantitative nature of child psychopathology

Rather than disease-like entities that are either present or absent, most childhood problems can be regarded as quantitative variations on behavior that is typical for children of a particular age and sex. With the possible exception of a few relatively rare conditions such as infantile autism or Tourette's disorder, most child psychiatric disorders do not consist of clearcut categories that can be easily separated from each other. Many otherwise normally functioning children may show deviant behavior to some degree at some time during their development. Age, degree of deviance, duration of the behavior, and, finally, the impact of the behavior on other areas of functioning must be considered in judging whether a child's development is in danger.

Population based data are needed to determine the degree to which a child's behavior deviates from that of same-sexed age-mates. As with pure somatic measures such as blood pressure or body temperature, the normative developmental approach to behavioral-emotional problems can impose cutoff points on distributions of scores to facilitate decisions. Cutoff points are usually chosen according to an external morbidity criterion, such as referral to a mental health service.

Lack of generally accepted assessment procedures

Child psychiatric research has been handicapped by a lack of generally accepted standardized assessment procedures. The use of different methodologies and different assessment instruments hampers comparisons between studies dealing with the same problems. In a field like child psychopathology where research activity is relatively scarce, it is especially essential to combine efforts in order to build up a common core of knowledge.

In an earlier report, we described how the prevalence of behavioral problems and social competencies reported by parents of Dutch children aged 4-16 (2) was assessed by using the Achenbach Child Behavior Checklist (CBCL) (4). The CBCL is designed to obtain parents' reports of a wide variety of children's behavioral and emotional problems and competencies. Employing the same general (epidemiological) methodology as Achenbach and Edelbrock (3), we obtained results strikingly similar to those obtained in the United States (2,5). This finding supports the applicability and generalizability of this instrument in the two countries.

The mutual relationship between epidemiology and the development and testing of assessment instruments is important for both fields. Adequate instruments are needed for assessing the degree and type of problem behaviors in large samples of children. On the other-hand, epidemiological methods are crucial for constructing and interpreting an instruments' measurements, for example by obtaining normative data against which the assessed variables can be tested.

Adults as key informants

Young children are less able to give reliable accounts of their behavior or emotions than are adolescents and adults. Although it is usually important to interview children to obtain a full understanding of their disorders, Rutter and Graham (6) found that a direct interview with the child added little to the information already available from parents and teachers. Furthermore, it is usually parents, teachers, or other adults who decide whether help should be sought.

Parents are usually the most important sources, because they can report on their child's behavior across many situations. Even if their judgment is affected by their relationship with the child, their perceptions may have valid implications for the child's long-term adaptation.

Next to parents, teachers are often the second-most important informants about a child's functioning. Although teachers may only see the child in the classroom and on the playground, these contexts may reveal difficulties in a child's functioning not evident elsewhere. Academic and social skills are likely to be related to later adult adaptive functioning. Furthermore, unlike parents, teachers have the opportunity to compare a particular child with large groups of peers. Concentration problems and social malfunctioning may be more evident at school than at home.

Situational specificity of child behavior and informant variance

Because different informants relate to children in different ways and in different contexts, we often find variations in children's behavior across different situations, as well as differences in the informants' judgments.

The issues of situational specificity of the child's behavior and the differences between reports provided by informants who know the child in different contexts may be inseparable, because different contexts (e.g. home, school, playground) usually involve different people from whom information is obtained (e.g. parents, teachers, peers). It has long been known that reports from parents and teachers tend to show small overlap (7,8). Only 7-20% of the children identified as disturbed by one type of informant is also identified by the other. The low agreement between reports by different informants has led to rejection of one or both of them as unreliable or invalid (8,9). Each source of information can be valid in its own way, however, and discrepancies between data from different sources may in-fact be as informative as agreements between them. For example, a child who is depressed and functioning poorly both at home and school could require a different approach than a child who is depressed at home but cheerful or aggressive at school.

Children's behavioral repertoire is much more variable than that of adults. Furthermore, children are more susceptible to environmental influences. On the other hand, information from adults can be coloured by a number of subjective factors too. Parents' judgments, for instance, may be influenced by their tolerance of the child's behavior, or by their own hostility, denial, or psychiatric conditions. Teachers' judgments may be coloured by the child's disruptiveness and academic performance, by the size of the class, and by personality characteristics of the teacher.

In a meta analytic study, Achenbach et al. (10) analyzed all available recent data on relations between different informants' reports on children's behavioral and emotional problems. They found a mean correlation of .60 between informants having similar relations to the child (e.g. pair of parents, teachers, observers), but .28 between informants having different relations to the child. Clinicians dealing with different, often conflicting, sets of data need to integrate these in order to form a picture of the child and to make decisions concerning advice or therapeutic approach. As yet, there are no systematic rules for weighting information available from different sources. These rules may differ with the kind of problem and with characteristics concerning the child and adult informants. For example, a 15-year-old depressed adolescent's own account of his emotional status, such as the reporting of suicidal ideas, may need to be weighted more heavily than parental reports of the boy's manifest behavior. On the other hand, a teacher's report of concentration problems and a drop in academic performance may be crucial for a child whose parents are divorced and who does not manifest signs of distress at home.

PREVIOUS STUDIES

In the introduction we have pointed out that teachers' reports may contribute to the overall picture of a child's functioning. Poor adaptive and academic functioning deserves attention.

Although not the only source of information on children's school functioning, teachers are key informants because they spend the most time with children in the school-setting and they are usually the best informed about a child's day-to-day behavior in the classroom or during recess. Furthermore, teachers are often the first to notice that a child's functioning needs to be evaluated.

The purpose of the following review is to discuss and compare existing teacher assessment procedures. As pointed out earlier, normative data are needed to determine the degree to which each child's behavior deviates from that reported for other children. We therefore focussed on studies using teacher assessment procedures in population based surveys, potentially providing normative data.

Next, the results of these studies are reviewed and compared. We will arbitrarily limit this review to studies carried out after 1960, because variations in children's behavior over time or variations in teachers' judgments may make data from earlier studies obsolete.

Assessment of children's behavioral-emotional functioning in school

Teachers' reports of children's behavioral-emotional problems can be obtained by having the teacher fill in a standardized rating form or by interviewing the teacher. Peer ratings and direct observations by trained observers are other ways to obtain a picture of a child's school functioning. Of these four procedures, standardized teacher rating forms have the advantage of being relatively unobtrusive and easy to apply. The period over which a child's behavior is observed can be relatively long compared with the much shorter periods over which direct observations are usually feasible.

Teacher rating scales for which data on reliability and validity are available and which have been used in population based surveys include: the Achenbach Teacher's Report Form (13); the Conners Teacher Rating Scale (14); the Quay-Peterson Behavior Problem Checklist (BPC) (15); the Bristol Social Adjustment Guides (BSAG) (16); and the Rutter Children's Behaviour Questionnaire, Teacher Form (17).

The fact that these instruments have been used in population based surveys implies that normative data have been obtained. However, normative data for different ages and genders have been reported only for the TRF, the Conners Teacher Rating Scale, and the BPC.

Scoring format

The various forms differ in their scoring of behavioral and emotional disorders. Rutter (17) used an a priori clinical framework in which symptoms were grouped according to two syndromes representing conduct disorder and emotional disorder. A totally different approach was employed by Achenbach and Edelbrock (13), who factor analyzed teachers' ratings of large representative samples of clinically referred children for both genders and age-groups 6-11 and 12-16. Referred samples were used to reflect the syndromes of problems manifested by children who are considered to need mental health services. Furthermore, the authors provided normative data for comparing each child's total score and scores on each empirically derived syndrome with those of same-sexed age-mates. The rigour of this approach in which careful attention was paid to representativeness and to variations in child psychopathology across different ages and for each sex, is not shared to the same extent by any of the other instruments.

For the Conners Teacher Rating Scale and for the BPC, empirically derived syndromes and norms are also available. Although empirically derived factors were recently reported for the BSAG scored on a normative sample (18), the instrument's data still have to be scored according to five predefined mutually exclusive core syndromes.

Reliability and validity

The usefulness of an instrument depends on its psychometric properties of which reliability and validity are the most important. Test-retest and inter-rater reliabilities have been reported for all instruments except the BSAG, for which internal consistency and one-year stability have been reported (16).

Test-retest reliabilities have usually been higher ($r = .80$ or higher) for most instruments than the interrater agreement (between $r = .50$ and $r = .60$).

The most widely used validity criterion has been discrimination between children referred for mental health services versus nonreferred children. However, the decision to seek help is largely determined by the parents of the child. Therefore, the accuracy with which an instrument classifies children according to referral status may be higher for parent administered instruments than for teacher administered instruments. The percentage correct classification for the TRF (75.2%) was found to be lower than that for the parent questionnaire, the CBCL (84.5%).

Another reflection of a measure's validity is its association with another instrument designed to capture the same features. For example, the correlations between scores on the TRF and the Conners' Revised Teacher Rating Scale were .90 for conduct disorder, .62 for hyperactivity and .76 for inattentive-passive behavior, whereas the correlation for the total score on both instruments was .85 (13).

According to Conners, the validity of his teacher rating scale was supported by evidence for its sensitivity to treatment effects. In a double blind study, the Conners scale showed significant differences in scores between a

placebo group and a group of learning and/or behavior disordered children treated with dextroamphetamine (14).

Of the instruments cited, the TRF (13) has the most solid quantitative background as it uses empirically derived syndromes based on clinically referred children and uses norms based on data from normative samples. Profiles standardized for different age groups and both genders are available. Furthermore, most TRF items are comparable to those on the parent questionnaire, the CBCL. Those were the main reasons why we chose the TRF to obtain ratings in the present study.

Most population based studies of teachers reported problems used teacher ratings to obtain prevalence rates on individual items or used the ratings to obtain an impression of children's overall functioning.

The prevalence estimates of specific symptoms are influenced by item wording and the scoring format of the assessment instrument. Unfortunately, differences across instruments hamper comparison of the results from different studies.

Population based surveys of behavioral-emotional problems

Table 1 summarizes 13 population based surveys using teachers to provide prevalence data either on overall psychological functioning or on specific behavioral-emotional problems.

Other studies using teacher data on behavioral-emotional problems have been carried out but are not summarized here for a number of reasons. Some have focussed only on one specific symptom (19,20), whereas others were confined to specific school settings (21). Many studies reported only the factor structures or other psychometric properties of the instrument (e.g. 22, 23, 24).

As can be seen from table 1, the sample size, age range, and method of sample selection show large variations across the different studies.

The American (25,26) and British (27,28) national surveys excel as far as sample representativeness is concerned. However, these studies reported few, if any, specific behavioral problems and used poorly validated definitions of maladjustment or psychiatric disorder.

In the British National Child Development Study, the sample of 16,000 children consisted of all those born in England, Scotland and Wales during a week in March 1958. Parent and teacher ratings of behavior problems were obtained at ages 7 (27) and 16 (28). In the US national surveys, data were collected on physical health as well as on the behavior of children at home and at school. For 7,417 6-11-year-olds and 6,768 12-17-year-olds data were collected.

Kellam et al. (29) reported rates of teachers' impression of maladjustment in 2,010 first-grade children in Woodlawn, Chicago, USA. The relatively high rate of maladjustment (33%) may partly be attributed to the low socio-economic status of Woodlawn, a mainly black, poor, urban neighborhood.

In an Australian study using the Connors Teacher Rating Scale in a pre-school and primary school population, disorder was defined on a purely statistical basis (30). Significant disorder was defined as 2 standard deviations above the sample mean and severe disorder as 3 standard deviations above the

sample mean. The author compared the 9.3% severe disorder in her study with the 6.8% psychiatric disorder found by Rutter et al. (31) in the Isle of Wight Study. However, as will be outlined below, the approach introduced by Rutter et al. (31) is so different from the one used in the Australian study that comparison may not be valid.

Rutter et al. (31) and Connell et al. (32) used a two-stage sampling procedure. Teacher- and parent questionnaires were administered to the target population in the first stage to identify possible problem children on the basis of their scores. The second stage involved a more intensive approach in which the previously identified children and their parents were interviewed and clinically judged.

This two-stage approach combines standardized and statistical procedures with clinical decision making processes. The final integration of information of the parent, teacher and clinician is left to the clinician. The advantage of this approach is that the end results are prevalence rates for different disorders generally accepted by practicing clinicians, which may be more appealing to them than scale scores or other statistically based rates. However, as we found when we used the same two-stage approach in an earlier published study (33), clinical judgment is rather arbitrary and may be strongly influenced by the investigators' clinical background. This makes comparisons between results from different research centers hazardous, even when agreement between raters within the same setting is satisfactory.

McGee et al. (11) used teacher and parent versions of the Rutter Questionnaire in a sample of 7-year-old children from Dunedin-New Zealand. In order to identify children with behavior problems, the authors used the cutoff scores used by Rutter et al. (31) for their 10-12-year-old sample. No allowance was made for age effects, which may account for the 31% of New Zealand children scoring above the cutoff, compared with only 12% in Rutter's study. No prevalence rates for specific symptoms were reported in the New Zealand study.

Sheperd et al. (34) obtained prevalence rates on 21 specific symptoms reported by teachers of a random sample of 6,463 children aged 5-15 attending local authority schools and special schools in Buckinghamshire, UK. The relatively small number of items omitted problems children often show in school, such as poor peer relations, and attentional problems.

Three studies reported specific item frequencies in entire school populations in different areas using the Quay-Peterson Behavior Problem Checklist (35, 36, 37). No comparisons between these studies were made.

Achenbach and Edelbrock (13) provided prevalence rates of problem symptoms and adaptive characteristics using the TRF in a random sample of 1,100 non-referred children aged 6-16 from public and parochial schools in the Omaha (Nebraska), Nashville (Tennessee) and Pittsburg (Pennsylvania) areas of the US.

In summary, this overview of teacher derived prevalence rates of children's behavioral-emotional problems reveals a lack of comparability across the most studies, due to the use of different instruments and different methodologies.

Psycho-educational assessment

Of the 14 population surveys listed in table 1, eight reported data on ability and/or achievement. Two (26,36) used global teacher impressions to rate children's psycho-educational functioning. Five (11,25,29,31,36) used IQ-(sub)tests and all but one (36) also reported achievement results. The study by Davie et al. (27) employed tests of reading, perceptual-motor functioning and arithmetic.

All studies that tested the relations between overall emotional adjustment and IQ, achievement or global teacher scores found them to be positively correlated. The U.S. National Health Survey (25), for example showed that only 13% of the well adjusted group of 12-17-year-olds had below average intellectual ability, whereas 55% of the maladjusted group had.

Rutter et al. (31) studied the relations between psychiatric disorder and psycho-educational functioning in greater detail. They found that IQ was slightly below average in antisocial boys, but not in antisocial girls, and in neurotic girls but not in neurotic boys. A much stronger association for both sexes was found between reading retardation and antisocial or mixed antisocial-neurotic disorders. As many as two-fifths of the children with a disorder involving antisocial symptoms were severely backward in reading. A cross-sectional study like this, however, cannot reveal causal relations. It is not clear whether antisocial behavior is caused by reading problems, whether reading problems are caused by antisocial behavior, or whether common factors such as temperamental characteristics underlie both behavioral and reading problems.

Factors associated with behavioral-emotional problems

Eight of the 14 studies listed in table 1 investigated the relation between behavioral-emotional problems and environmental factors.

Problem behaviors were more prevalent in lower socio-economic status children in 3 of the 4 studies reporting on this effect (13,27,31,34). Only Rutter et al. (31) found no clear SES-effect on the prevalence of psychiatric disorders, whereas Sheperd et al. (34) found a significant SES-effect only in girls.

Black children showed slightly, though significantly, more teacher reported problems than white children (13). Children from schools in lower SES-catchment areas showed more behavioral-emotional problems than children from higher SES-catchment areas (30,36).

Davie et al. (27) found that mother's smoking during pregnancy, too early or too late birth, and large family size were positively associated with poor social adjustment. Rutter et al. (31) found that antisocial children tended to come from large families and neurotic children from small families. Furthermore, they found that childhood psychiatric disorders were more prevalent in homes broken by the death, divorce or separation of the parents.

Both the U.S. National Health Surveys (25,26) found that poorly adjusted children were less well accepted and perceived as less popular by their peers

than were well adjusted children.

In conclusion, children showing behavioral-emotional problems were found to be disadvantaged in a number of psycho-educational and social areas as well.

Table 1 Population - based surveys using teachers as informant

| Authors, yr. of publication Study Location | Sample | | Definition of Disorder | Assessment Method(s) | Prevalence of psychiatric disorder | | | Intellectual Ability; Academic performance | Associated Factors |
|--|--------|--------------|---|--|--|--|--|--|---|
| | size | age | | | method | overall ¹ | specific syndromes | | |
| Achenbach et al. 1986 U.S.A. (13) | 1,100 | 6-16 | random selection of children from public and parochial schools | Teacher Report Form (TRF) | n.a. ² | 33 empirical syndromes for both sexes and ages 6-11, 12-16 | 118 symptoms reported | n.a. | socio- economic class |
| Connel et al. 1982 Queensland, Australia (32) | 366 | 10-11 | random sample of schools; two-stage sampling procedure | Rutter Children's Behaviour Questionnaire for completion by parents and teachers; parent and child interview in second stage | 10% psychiatric disorder | n.a. | n.g. ³ | n.a. | n.a. |
| Davie et al. 1972 Pringle et al. 1966 U.K. (27) | 15,039 | 7 | whole sample of birth cohort using census data | British Social Adjustment Guides (BSAG) educational and medical assess- ment; parent questionnaire | 14% maladjusted | 10 teacher syndromes reported | no teacher items; 23 parent symptoms reported | reading, percentual- motor and arithmetic ability tests | biological factors; social factors |
| Fogelman 1976 U.K. (28) | 12,264 | 16 | whole sample of birth cohort using census data | Rutter Children's Behaviour Questionnaire for completion by parents and teachers | n.a. | n.a. | 18 symptoms reported | n.a. | n.a. |
| GLOW 1976 Aelaide, Australia (30) | 2,475 | grade k-6 | random sample of schools; urban | Connors Teacher Rating Scale | 29% significant severe teacher perceived disorder | 7 empirically derived teacher syndromes | 39 symptoms reported | n.a. | school and teacher variables |

| | | | | | | | | | | |
|--|-------|---|--|--|---|--|--|--|---|---------------------------------|
| Kellam et al. 1975 Chicago Illinois U.S.A. (29) | 2,010 | grade 1 and low- up at grade 3 | whole population first graders; poor, urban | global teacher's perception maladap- tation | own questionnaires for teachers, observers, selfratings | 19.4% moderately and 13.6% severely maladapted | n.a. | 5 teacher symptoms, 5 observer; 7 selfratings, 38 parent symptoms | IQ and reading achievement tests | n.a. |
| Mc Gee et al. 1984 Dunedin, New Zealand (11) | 951 | 7 | total sample of children born in one hospital | children whose scores fell at or above predefined cutoff score | Rutter Children's Behaviour Questionnaire for parents and teachers; behavior ratings during psychometric assessment | 30% children with high level of behavior problems | 3 broad categories: neurotic, antisocial, undifferen- tiated | n.g. | IQ and reading ability tests | n.a. |
| Oliver 1974 U.S.A. (25) | 6,768 | 12-17 | random sample general population; national | global teacher's perception of emotional adjustment | own teacher questionnaire | 18.7% boys and 12.3% girls maladjusted; 75% well- adjusted | n.a. | n.g. | ability and performance ratings by teacher; subtests of standardized achievement test | peer popularity |
| Roberts et al. 1972 U.S.A. (26) | 7,417 | 6-11 | random sample general population; national | global teacher's perception of emotional adjustment | own teacher questionnaire | 23.7% boys and 9.6% girls maladjusted | n.a. | 8 symptoms reported | ability and performance ratings by teacher | peer acceptance |
| Rutter et al. 1970 Isle of Wight, U.K. (31) | 3,316 | 10-12 | total population of children attending public schools and children in certain admini- strative groups; two- stage sampling procedure | clinical diagnostic | Rutter Children's Behaviour Questionnaire for completion by parents and teachers | 6.8% psychiatric disorder, excluding uncom- plicated mental retardation and mono- symptomatic disorders | 16 syndromes reported | 26 teacher symptoms, 31 parent symptoms | IQ and reading ability tests | family and social factors |

| Author | Year | Sample Size | Grade | Population | Prevalence Item | Questionnaire | Notes | Symptoms Reported | Outcome |
|----------------------------------|------|-------------|-----------|---|-----------------|---|--|----------------------|---|
| Schultz et al. | 1974 | 1,575 | 3 and 4 | total population of elementary school children in two countries | item prevalence | Behavior Problem Questionnaire | n.a. | 55 symptoms reported | n.a. |
| Essi-cenral Illinois U.S.A. (35) | | | | | | | | | n.a. |
| Sheperd et al. | 1971 | 6,463 | 5-15 | one-in-ten sample of children attending all local authority schools | item prevalence | own teacher and parent questionnaire | 55% boys and 61% girls had no behavior problems noted by teacher | 21 symptoms reported | general attainment and reading ability rated by teacher |
| Buckinghamshire, U.K. (34) | | | | | | | | | socio-economic status |
| Stone 1981 | 1981 | 24,997 | grade k-6 | entire school population of one town | item prevalence | Behavior Problem Questionnaire, shortened version | n.a. | 36 symptoms reported | intelligence test scores for grades 4 and 6 |
| Des Moines, Iowa U.S.A. (36) | | | | | | | 2 syndromes: conduct and personality problems | | n.a. |
| Werry et al. | 1971 | 1,763 | grade k-2 | entire school population | item prevalence | Behavior Problem Questionnaire | n.a. | 55 symptoms reported | n.a. |
| Illinois, U.S.A. (37) | | | | | | | | | n.a. |

1 overall = overall psychological functioning

2 n.a. = not assessed

3 n.g. = not given

AIMS OF THE PRESENT STUDY

It is clear that we need data from different observers (parents, teachers, clinicians, and children themselves), which, even if they disagree, may independently contribute valuable information for clinical decisions. At this point, efforts are needed to refine and test assessment tools to make each of the different pictures as valid as possible.

In an earlier report, we provided prevalence rates and cross-cultural comparisons of behavioral-emotional problems in Dutch children as reported by their parents (2). We also provided evidence for the reliability and validity of the instruments used, the Achenbach Child Behavior Checklist (CBCL) (4).

At the time the present study was started, there were no data available in the Netherlands on the prevalence of a wide range of teacher reported behavioral-emotional problems, nor were there instruments available for determining the degree and type of these problems.

Municipal birth registers in the Netherlands offer possibilities for obtaining prevalence rates and effects of demographic variables that reliably reflect those occurring in the general population. As will be shown in the literature review, most previous prevalence studies of teachers as informants used samples of convenience, or samples limited to particular populations, whereas others reported on only a narrow range of ages or behaviors.

A strong test of the generality of child psychopathological features is the cross-national replicability of results using the same methods and instruments. Except for studies comparing prevalence rates for different ages (11) or studies focussing on only one particular symptom such as hyperactivity (12), no systematic comparisons have been made between prevalence rates obtained from teachers' reports on the same instrument in different countries.

In the present study, prevalence rates for behavioral-emotional problems of children aged 4-12 as reported by their teachers were obtained using the Achenbach Teacher Report Form (13).

The aims of the study were:

1. To provide prevalence rates on a wide range of specific behavioral and emotional problems as reported by teachers in a representative sample of Dutch children aged 4-12 using the Teacher Report Form (TRF) (13);
2. To identify differences related to demographic variables;
3. To compare Dutch data with data from other population based surveys.

METHOD

Data collection procedure

The assessment instrument used was the Teacher's Report Form developed by Achenbach and Edelbrock (13) (see Appendix A). The TRF is a questionnaire designed to obtain Teachers' reports of children's behavioral-emotional problems and adaptive functioning in a standardized format. The TRF problem items have the same format as the Child Behavior Checklist (CBCL) (4), which is designed to obtain parents' reports of their children's problems and competencies. The CBCL was used in our earlier reported population survey (2) to obtain parent reported prevalence rates for the same sample on which our teacher data were obtained.

The TRF contains two parts: the first section has questions on academic performance and general adaptive characteristics of the child, whereas the second part includes 118 specific behavioral-emotional problems, plus open-ended items for adding physical problems without known medical cause and other problems not specifically listed.

The TRF was translated into Dutch with the help of a linguist. Efforts were made to refer as precisely as possible to the behavior covered by the original version.

Because the questions on academic functioning and on general adaptive characteristics were scored somewhat differently in our study, comparison with the American data of these items could not be performed. We therefore report here only the result of the problem items.

Although the parent and teacher versions of the questionnaire are very similar, a number of items relevant to the one situation but not to the other, show differences. Of the 118 items, 25 such as "bedwetting" or "disobedient at home" are replaced by items more relevant to the school situation, such as "disrupts class discipline" and "inattentive, easily distracted".

The teacher is asked to rate the problem items on a 0-1-2 scale, based on the preceding two months. The teacher is asked to circle a 0 if the item is not true of the child; a 1 if the item is somewhat or sometimes true; and a 2 if the item is very true or often true. On a number of items the teacher is asked to describe the child's behavior, making it possible to correct the scoring when the described behavior does not fit the items.

All checklists were checked for inappropriate scoring and were corrected if necessary. When in doubt, we contacted the teacher.

In order to assess test-retest reliability, 44 checklists were mailed after a 3-4 week interval to randomly selected teachers. A reminder was sent to those teachers who did not return the TRF. Because the summer holidays interfered, we were not able to seek a higher response rate without varying the test-retest period.

Twenty-two TRF's were used for computing the test-retest Intraclass Correlation Coefficient (ICC), which was .84 for the sum of problem scores. Achenbach and Edelbrock (13) found a Pearson r of .84 for an interval of 15 days and an r of .90 for an interval of 7 days.

Description of sample

The sample was drawn between February and May, 1983, from Zuid-Holland, a province in The Netherlands encompassing over 3,000,000 people living in urban, semi-rural and rural areas.

In 1981, there were 572,238 4-16-year old children of Dutch nationality in Zuid-Holland (51% boys; 49% girls) (38).

At the time of the survey, compulsory education encompassed ages 6-16. In addition, about 95% of all 4-year-olds and 98% 5-year-olds attended kindergarten (39). About 6% of children aged 6-12 attended special schools. Most elementary schools and kindergarten are within 15 minutes walk (or bicycling) of the child's home. Classroom size is more than 30 for over 50% of the schools, whereas only 25% have classrooms with less than 25 children.

In an earlier report on our population survey of parent-reported problems in children aged 4-16 we described the sampling and interviewing procedure in detail (2). Using municipal birth registers that list all residents, we drew a random sample of 100 children of each age and sex with the Dutch nationality (total N = 2,600). Two of the selected municipalities declined to participate. One municipality first contacted the parents of selected children to request permission in advance, five of whom, declined to participate. Of the 2,447 parents reached, 2,076 (84.8%) completed the interviews.

All parents of children attending kindergarten and elementary school were asked by the interviewer to give written permission for their child's teacher to fill in the TRF. We sought TRF's only for kindergarten and elementary school-children because secondary school teachers are usually less well informed about their pupils. By including only kindergarten and elementary school-children, we focused on a teacher-child relationship that was comparable across different age groups.

TRF's were initially mailed to the teachers in April, 1983. At the end of June, just prior to the summer holidays, a reminder was sent to those who had not returned the complete questionnaire. In this way, information was obtained from teachers who had known the child for a minimum of eight months.

Completed TRF's were obtained on 1,162 children (79.2%) aged 4-12. Table 2 gives the distribution of the sample by sex and age.

For some statistical analyses and for comparison with the American sample, a normative sample was composed by excluding 19 children referred to a mental health agency, 36 children attending schools for special education, and 3 children receiving both types of help.

Socio-economic status (SES) was scored on a six-step scale of parental occupation (40). If both parents worked, the higher-status occupation was used. Table 3 shows the percent of each occupational level for the parents in the general population sample. The mean SES score was 3.58 sd = 1.55, which was slightly above the midpoint of 3.5 (6 = highest status occupation).

Because parental information on the child's behavioral-emotional problems was available for those children on whom no teacher data were obtained, we could determine whether the level of parent-reported problems in the sample for which both teacher and parent data were available differed from the one for which teacher data were lacking. The mean total behavior problem-score on

the parent checklist of 22.14, sd = 16.0 for the sample having both teacher and parent data did not differ significantly from the total score of 22.09, sd = 17.09 of the sample without teacher information.

Referred sample

In order to assess the discriminative power of the TRF, we used a sample of 57 children aged 4-11 referred to our outpatient clinic. The small sample size and the fact that it was obtained from only one clinic limits the representativeness of this sample. However, by matching this sample for age, gender and SES with an equally sized sample of nonreferred children, we explored the sensitivity and specificity of the TRF as well as the level of association of the problem items to referral status.

Table 2

Distribution by age and gender of general population sample for whom TRFs were obtained

| Age (yrs) | General population sample | | Normal sample | |
|--------------|---------------------------|-------|---------------|-------|
| | Boys | Girls | Boys | Girls |
| 4 | 45 | 53 | 44 | 52 |
| 5 | 72 | 79 | 70 | 78 |
| 6 | 62 | 71 | 60 | 71 |
| 7 | 57 | 75 | 50 | 72 |
| 8 | 74 | 63 | 73 | 58 |
| 9 | 63 | 67 | 54 | 65 |
| 10 | 61 | 69 | 56 | 67 |
| 11 | 64 | 81 | 58 | 77 |
| 12 | 55 | 51 | 49 | 50 |
| Total | 553 | 609 | 514 | 590 |

Table 3

Distribution of occupational level of parents of children in the general population sample.

| Occupational level | General population sample N = 1162 % |
|---|--|
| 1. Unskilled employees | 5 |
| 2. Skilled manual employees | 28 |
| 3. Clerical, technicians, minor professionals | 22 |
| 4. Owners of small businesses | 12 |
| 5. Supervisory, lesser professionals | 17 |
| 6. Executives, major professionals, owners of large businesses | 16 |

RESULTS

Prevalence of specific behavioral-emotional problems

For each specific problem item, the percentage of children grouped by age and gender for whom the behavior was reported by the teacher is graphically shown in figures 1-112. To make our data comparable to Achenbach and Edelbrock's (13), children were grouped by age intervals 4-5, 6-7, 8-9, 10-11 and 12. Because behavioral-emotional problems have been found to vary with SES, the percentages depicted in the figures were standardized for SES in tertiles. Scores of 1 and 2 were combined to provide the percentages.

Although the points in figures 1-112 are connected for reasons of clarity, this does not imply that the development of an individual child's behavior will follow the outlined pattern.

In each figure the effect of age and gender on the prevalence rates is indicated. The graphs give percentages of children for whom the problem was reported by combining scores of 1 and 2. This was done for reasons of simplicity of presentation. However, for statistical analysis, the original 0-1-2 scoring was retained.

Figures 1-112 Percentage of children in the general population sample (N = 1,162) of each gender for whom each behavioral or emotional problem was reported by the teacher. Scores of 1 and 2 for each item are combined to obtain the percentage for whom the problem was reported. For statistical analysis of sex and age differences that are reported in the graphs, the original 0-1-2 scoring was retained.

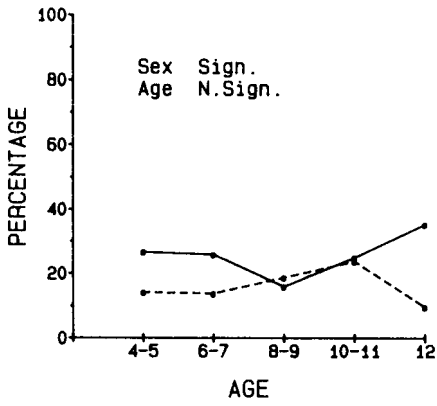
Figure 113 Mean total problem scores.

----- Girls
_____ Boys

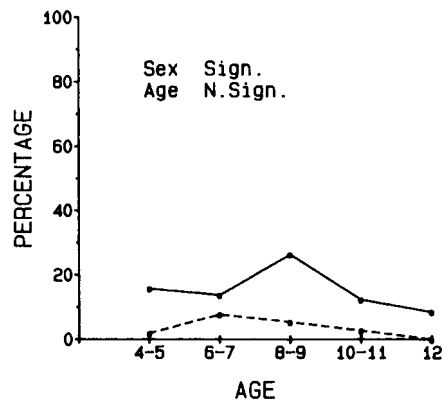
Sign. = Significant effect ($p < .01$) in ANCOVAs.

N.Sign. = Non-significant effect in ANCOVAs.

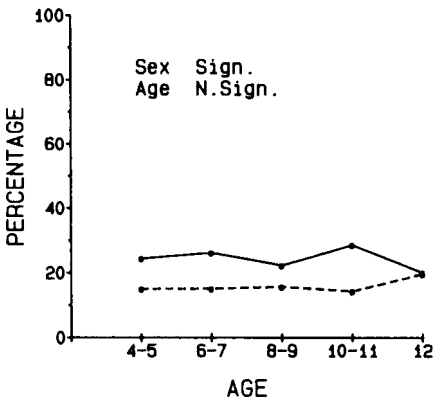
1. ACTS YOUNG



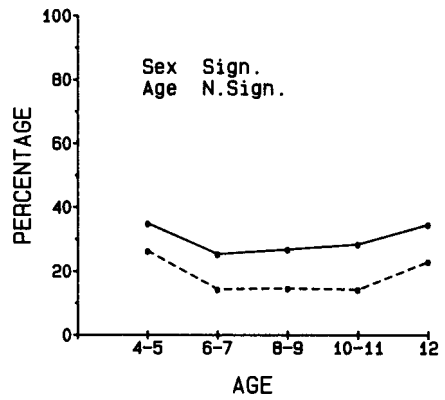
2. HUMS



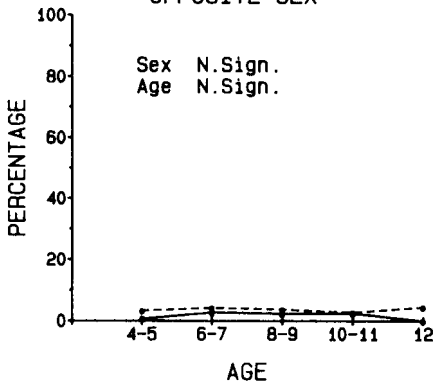
3. ARGUES



4. FAILS TO FINISH



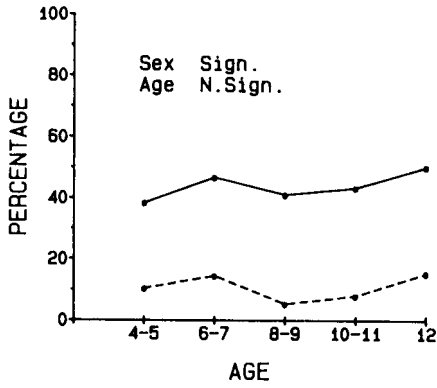
5. BEHAVES LIKE OPPOSITE SEX



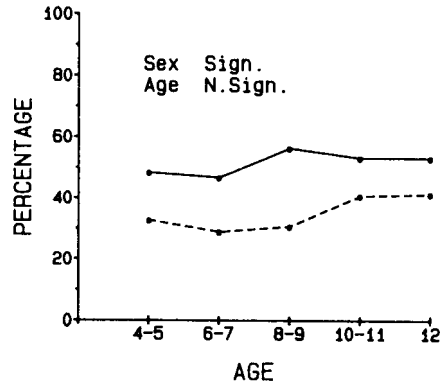
6. DEFIANT



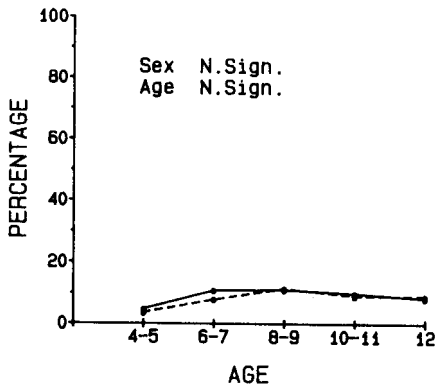
7. BRAGGING



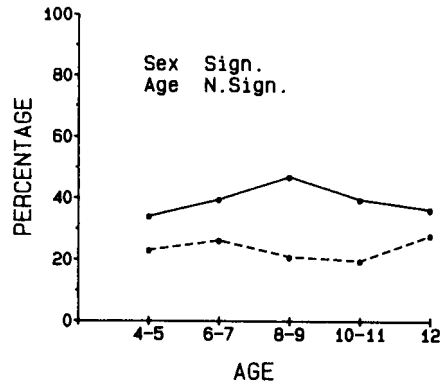
8. CAN'T CONCENTRATE



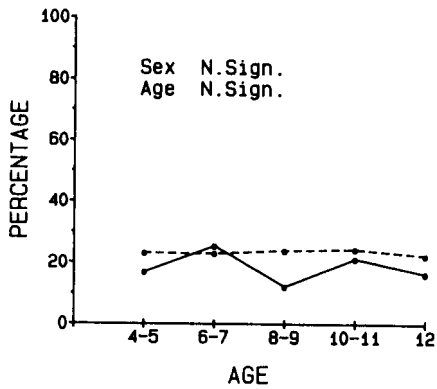
9. OBSESSIONS



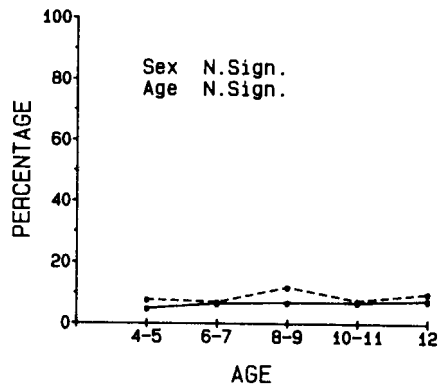
10. HYPERACTIVE



11. TOO DEPENDENT



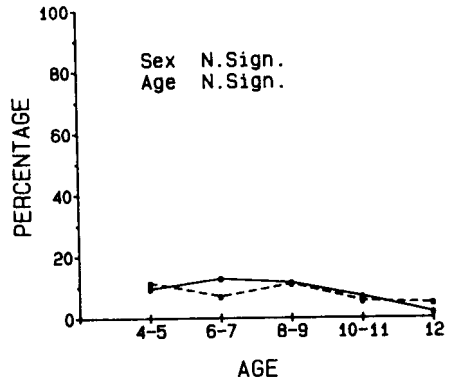
12. LONELY



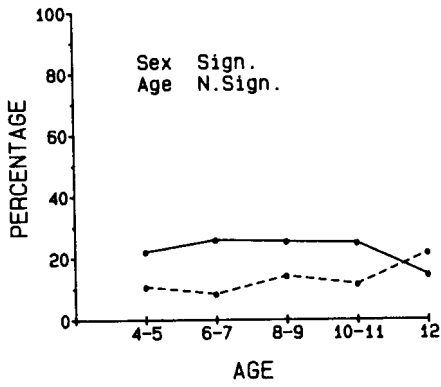
13. CONFUSED



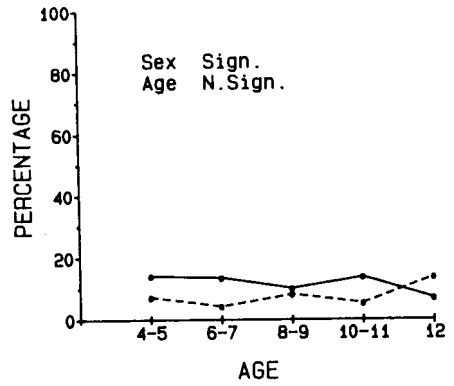
14. CRIES



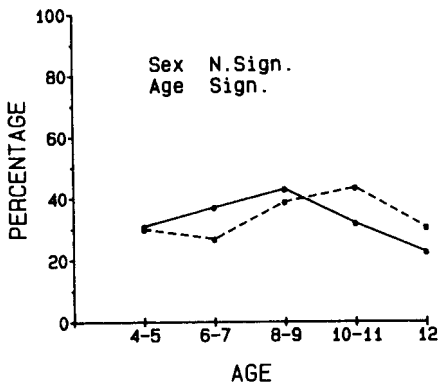
15. FIDGETS



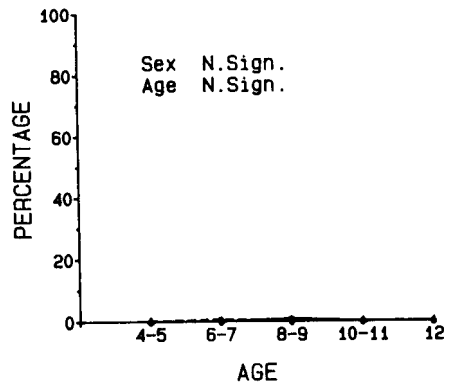
16. CRUEL TO OTHERS



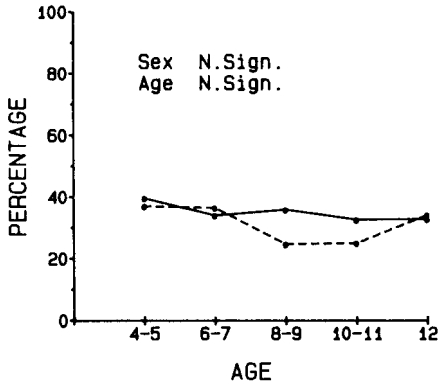
17. DAY-DREAMS



18. HARMS SELF



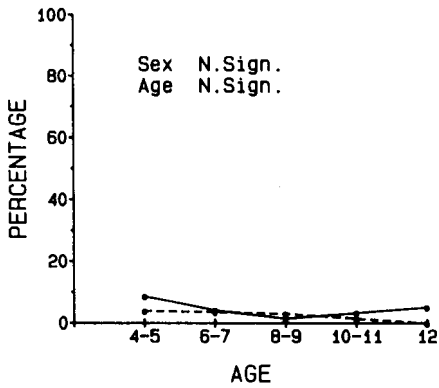
19. DEMANDS ATTENTION



20. DESTROYS OWN THINGS



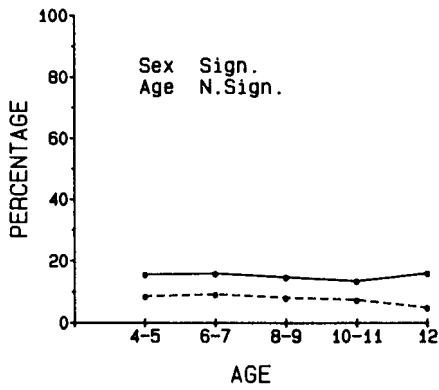
21. DESTROYS OTHERS' THINGS



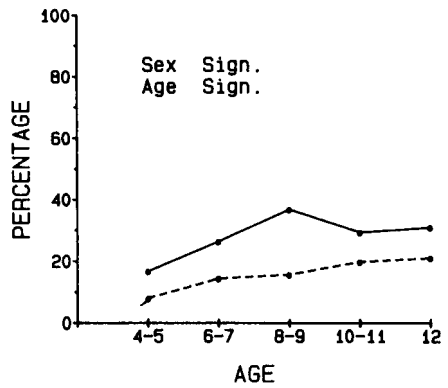
22. DIFFICULTY WITH DIRECTIONS



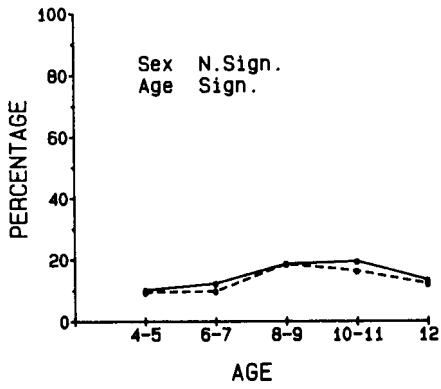
23. DISOBEDIENT



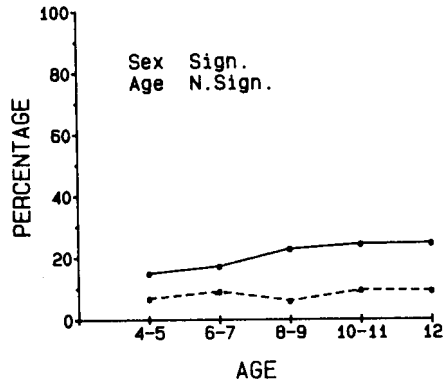
24. DISTURBS OTHERS



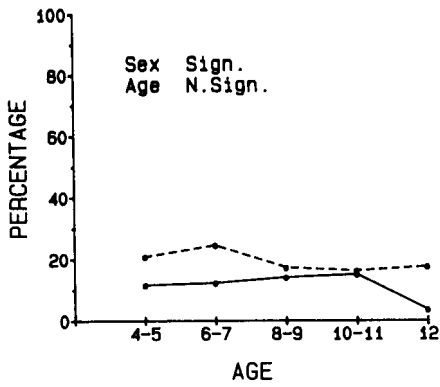
25. POOR PEER RELATIONS



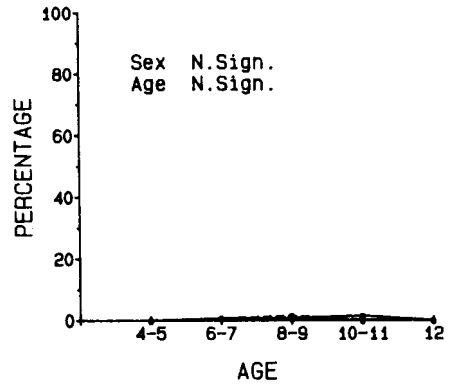
26. LACKS GUILT



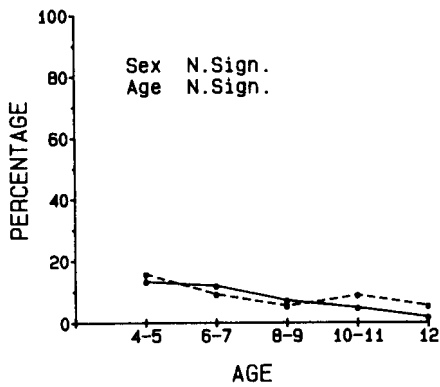
27. EASILY JEALOUS



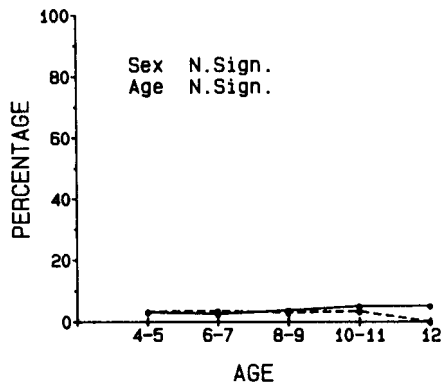
28. EATS NON-FOOD



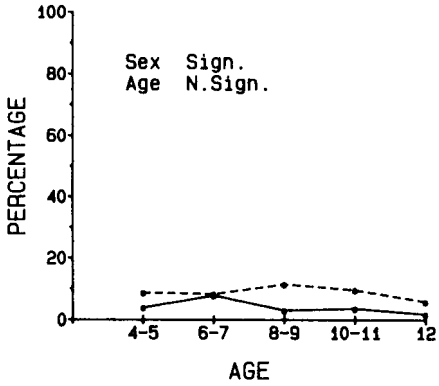
29. FEARS



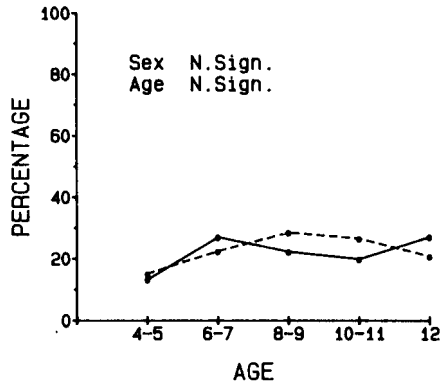
30. FEARS SCHOOL



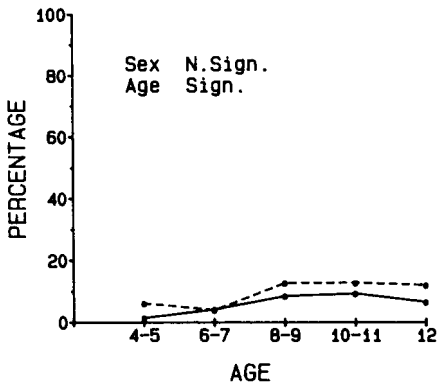
31. FEARS IMPULSES



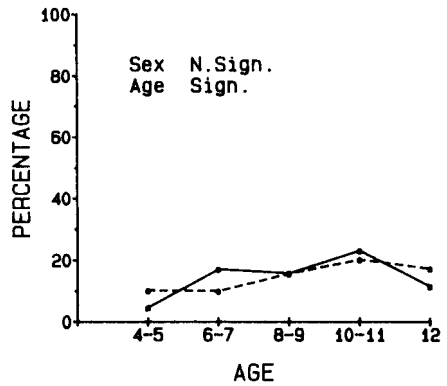
32. NEEDS TO BE PERFECT



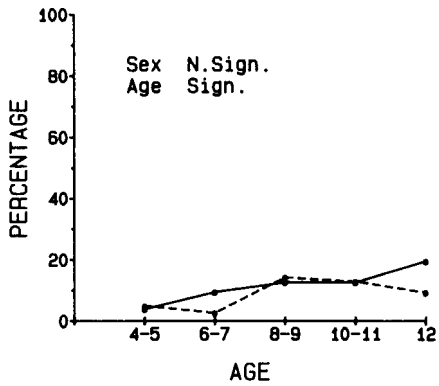
33. FEELS UNLOVED



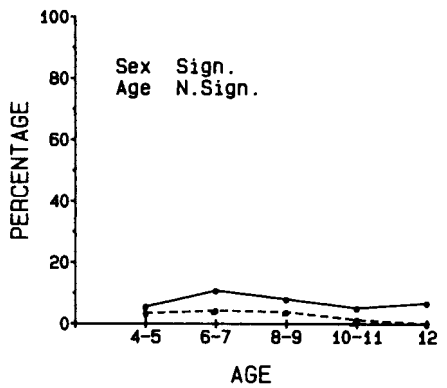
34. FEELS PERSECUTED



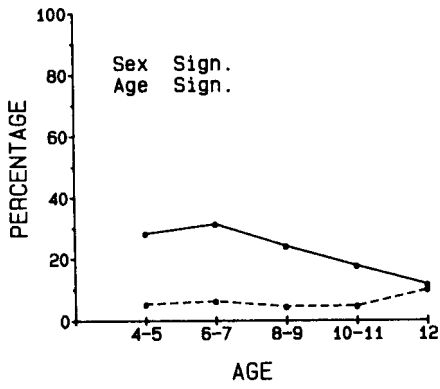
35. FEELS WORTHLESS



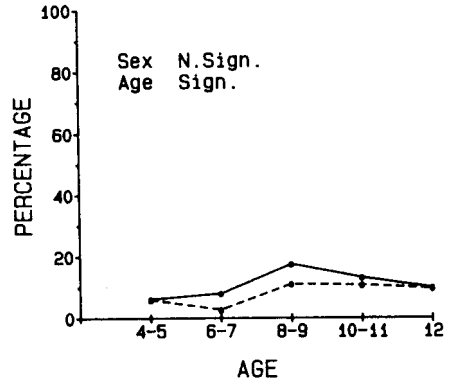
36. ACCIDENT-PRONE



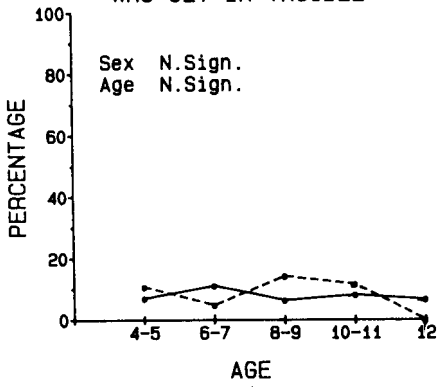
37. FIGHTING



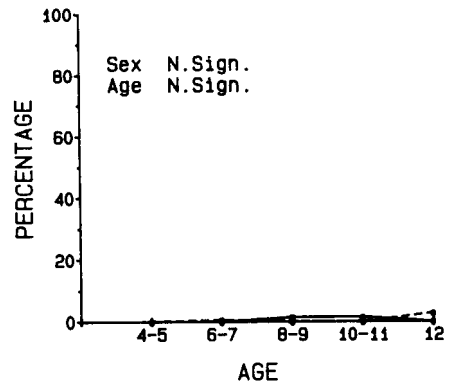
38. IS TEASED



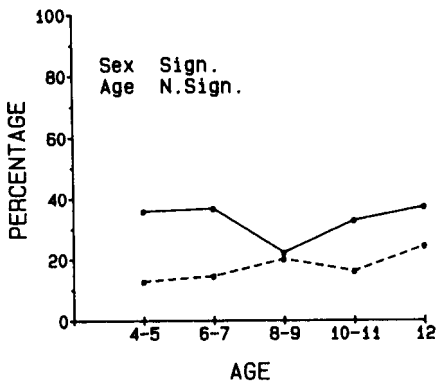
39. HANGS AROUND OTHERS WHO GET IN TROUBLE



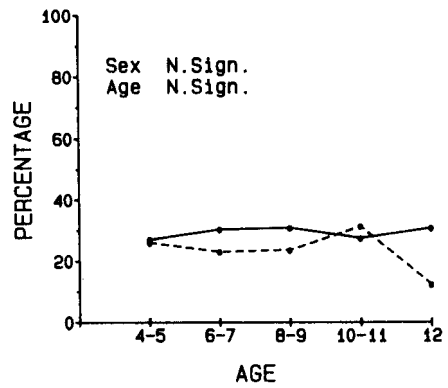
40. HEARS THINGS



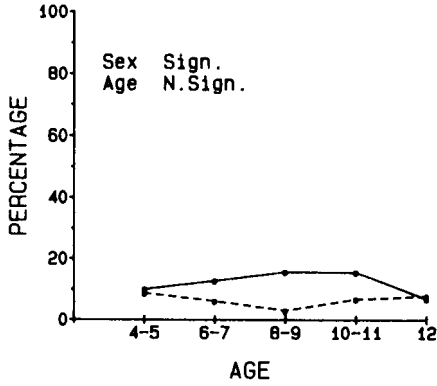
41. IMPULSIVE



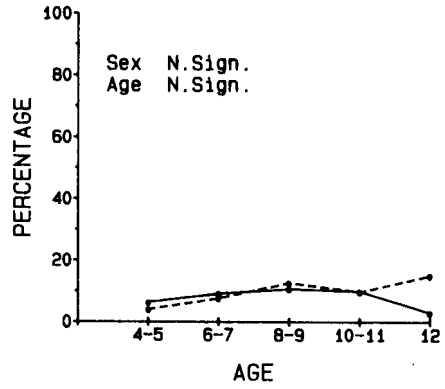
42. LIKES TO BE ALONE



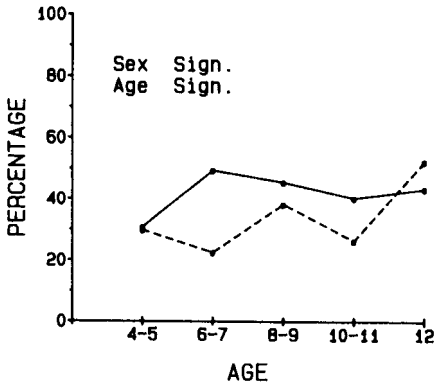
43. LYING OR CHEATING



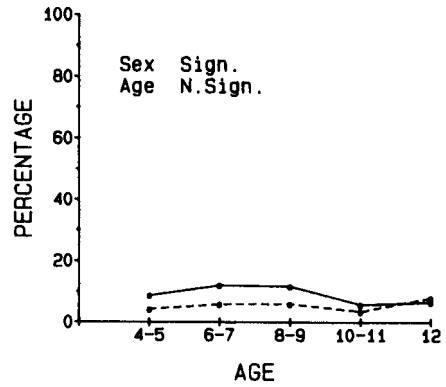
44. BITES FINGERNAILS



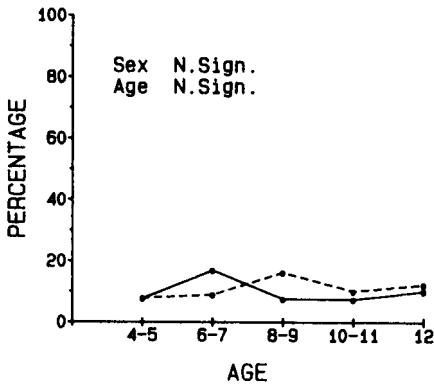
45. NERVOUS



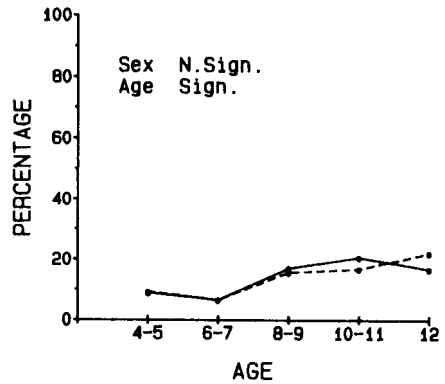
46. NERVOUS MOVEMENTS



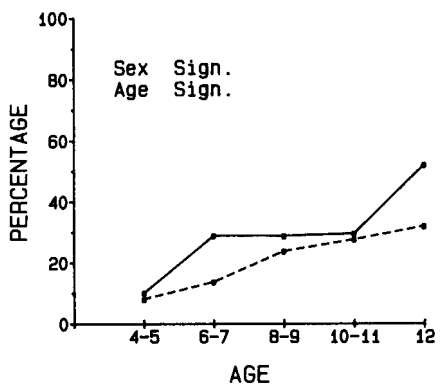
47. OVERCONFORMS



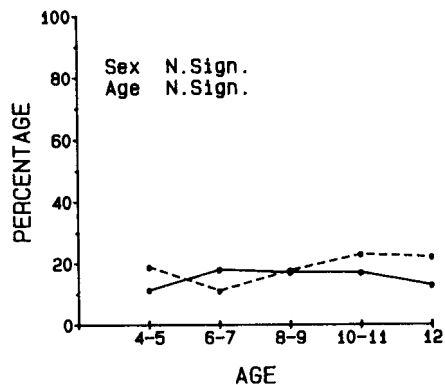
48. NOT LIKED



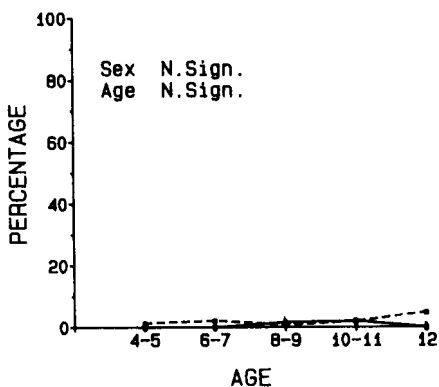
49. DIFFICULTY LEARNING



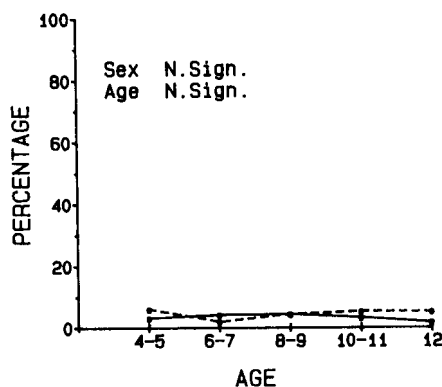
50. FEARFUL OR ANXIOUS



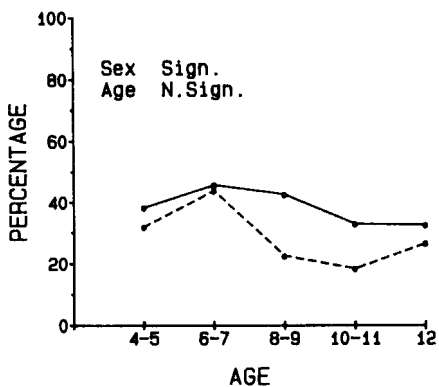
51. DIZZY



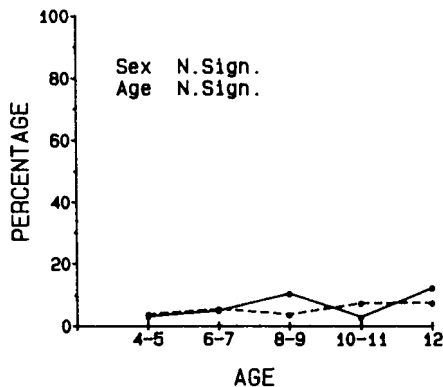
52. FEELS TOO GUILTY



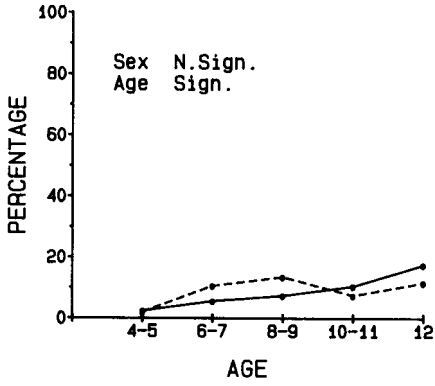
53. TALKS OUT OF TURN



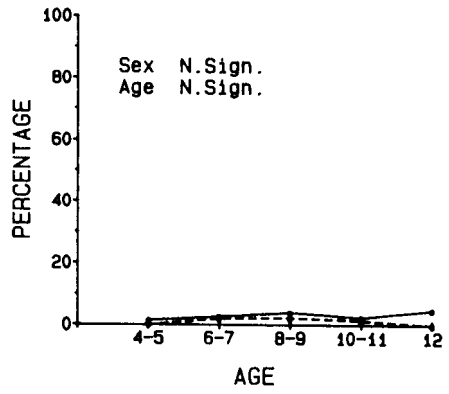
54. OVERTIRED



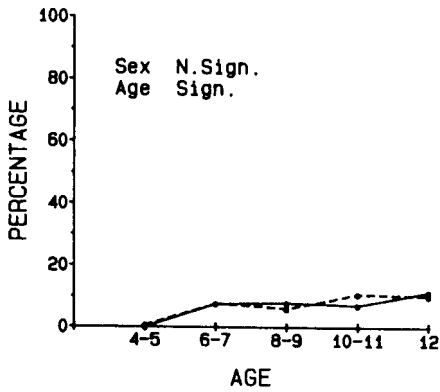
55. OVERWEIGHT



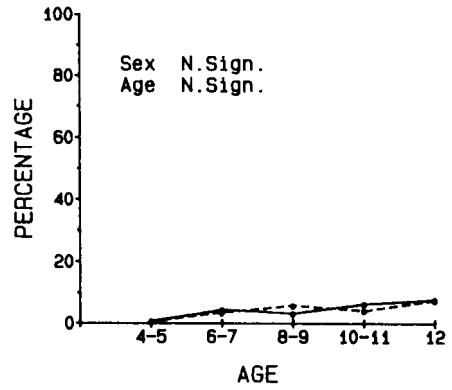
56A. ACHES OR PAINS



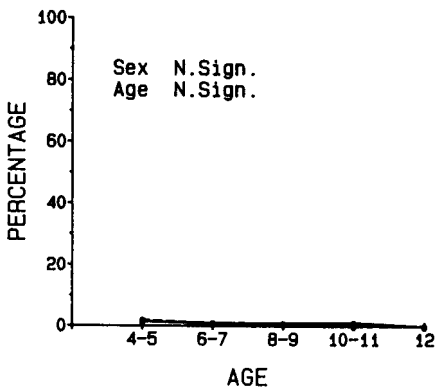
56B. HEADACHES



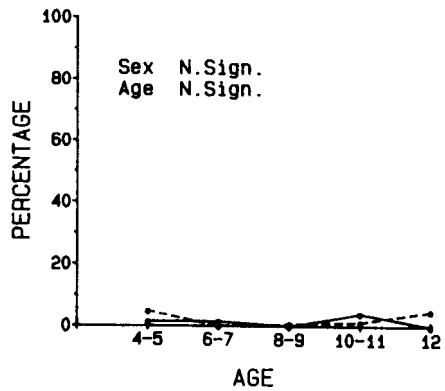
56C. NAUSEA, FEELS SICK



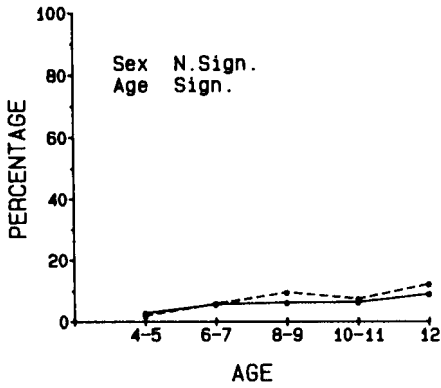
56D. EYE PROBLEMS



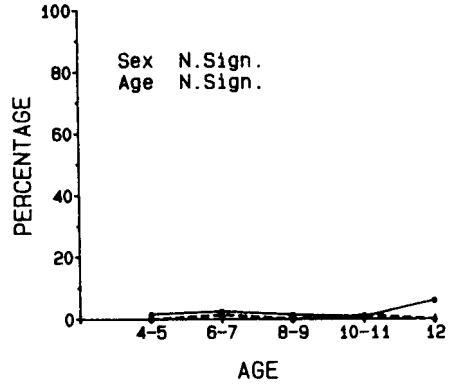
56E. SKIN PROBLEMS



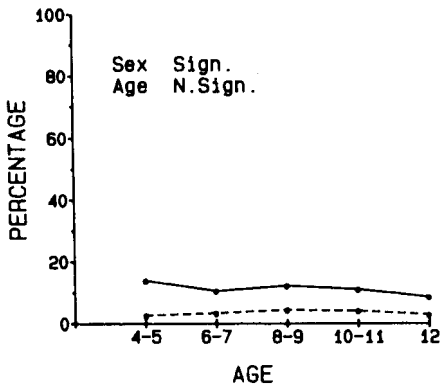
56F. STOMACHACHES



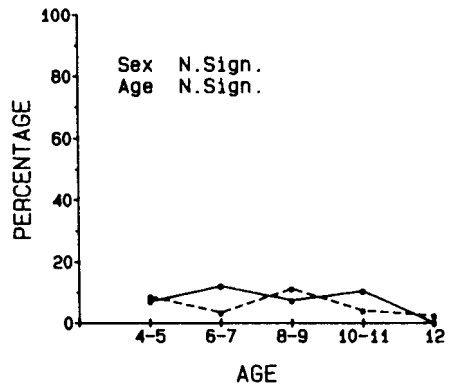
56G. VOMITING



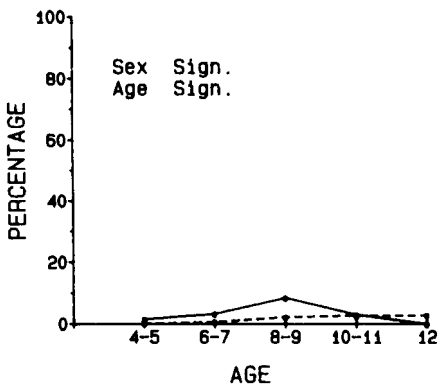
57. ATTACKS PEOPLE



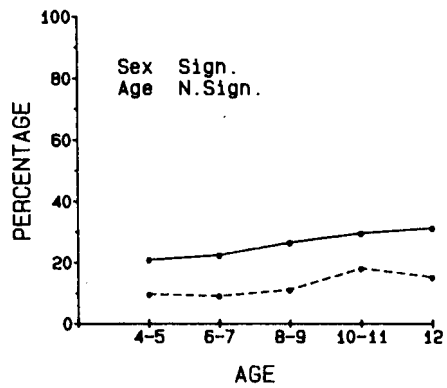
58. PICKING



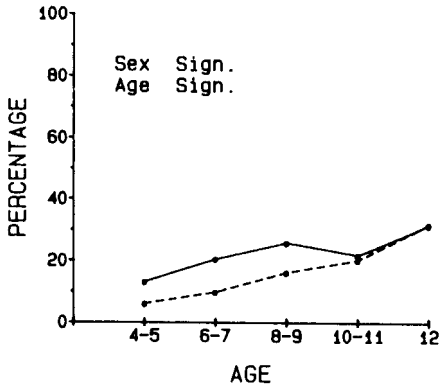
59. SLEEPS IN CLASS



60. APATHETIC



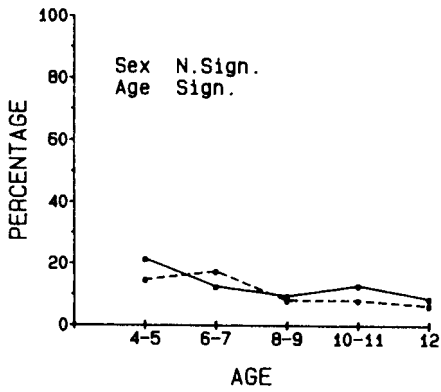
61. POOR SCHOOL WORK



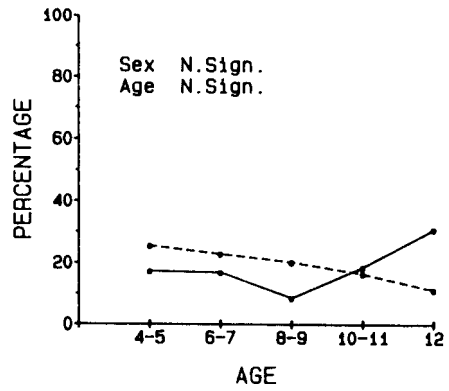
62. CLUMSY



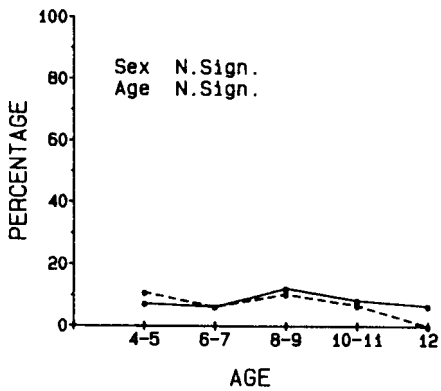
63. PREFERS OLDER CHILDREN



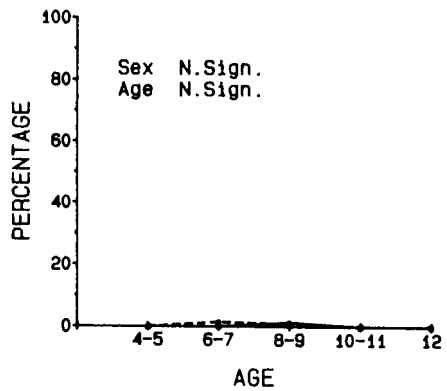
64. PREFERS YOUNGER CHILDREN



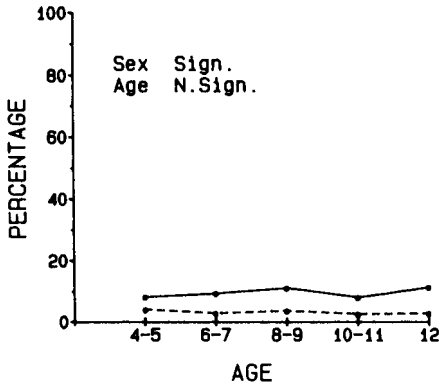
65. REFUSES TO TALK



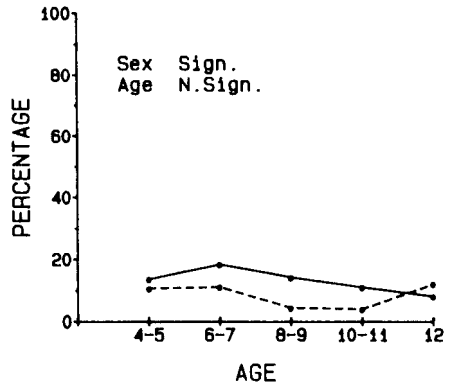
66. COMPULSIONS



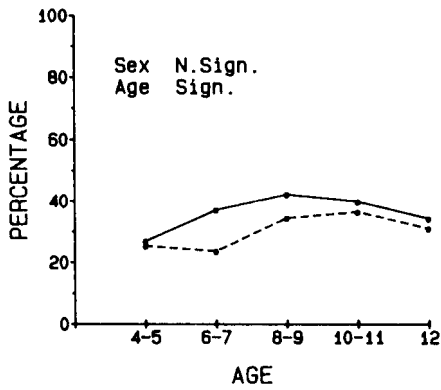
67. DISRUPTS CLASS



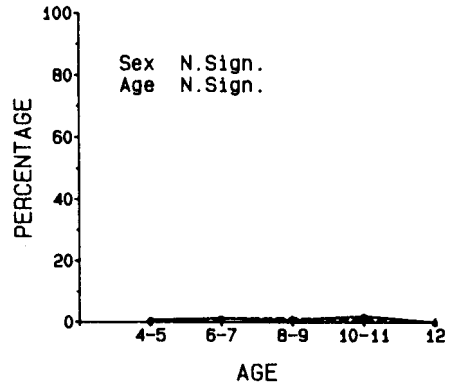
68. SCREAMS A LOT



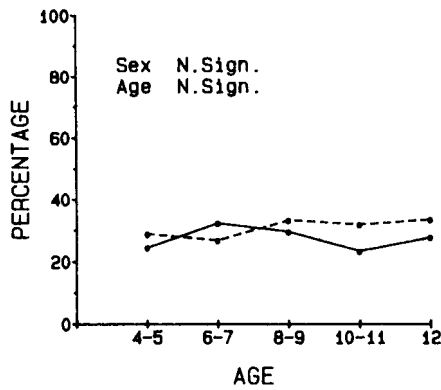
69. SECRETIVE



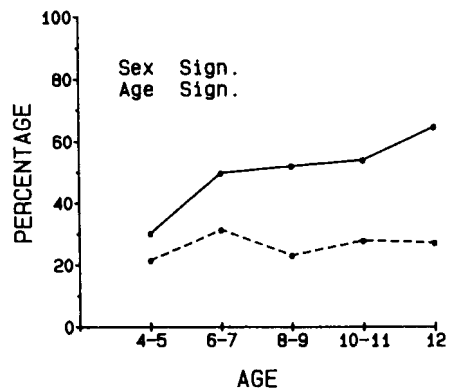
70. SEES THINGS



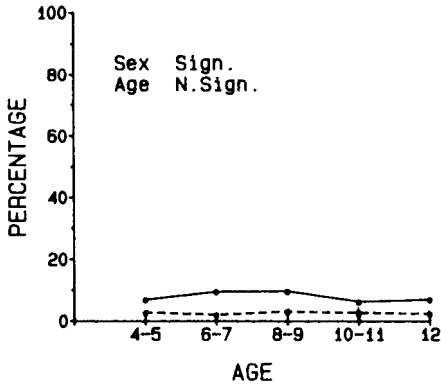
71. SELF-CONSCIOUS



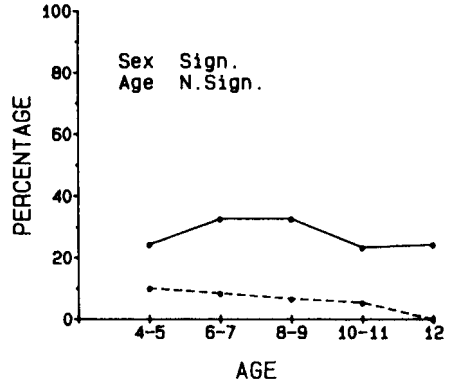
72. MESSY WORK



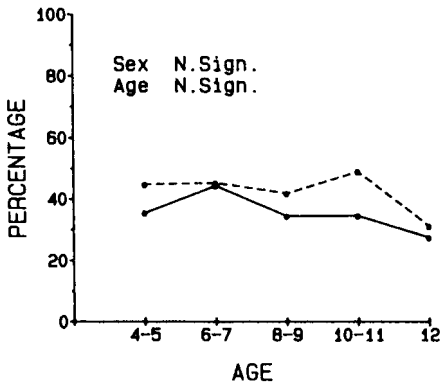
73. IRRESPONSIBLE



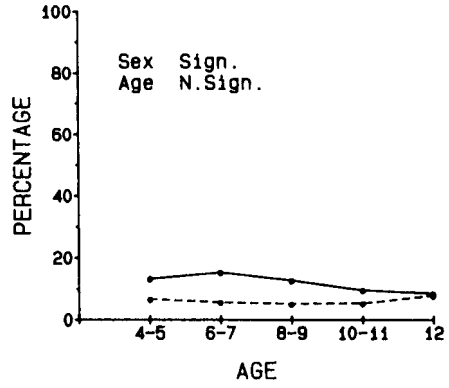
74. SHOWING OFF



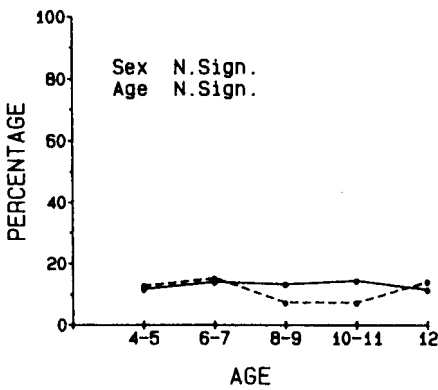
75. SHY OR TIMID



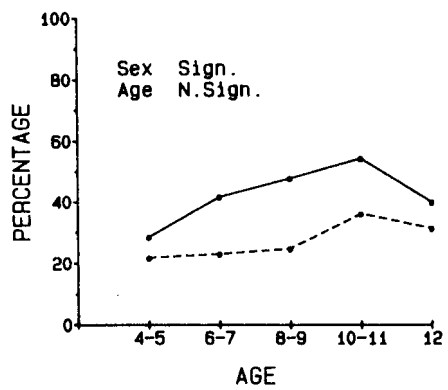
76. EXPLOSIVE



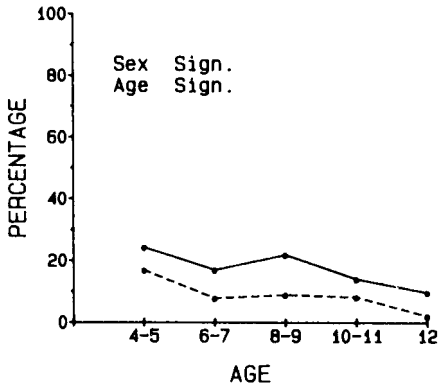
77. EASILY FRUSTRATED



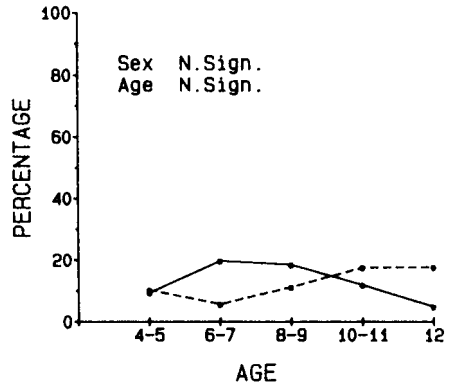
78. INATTENTIVE



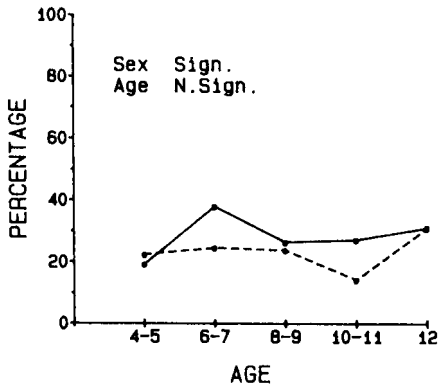
79. SPEECH PROBLEM



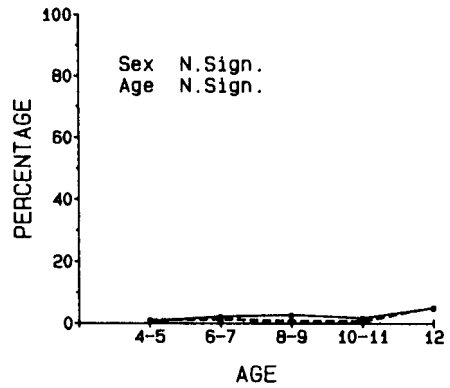
80. STARES BLANKLY



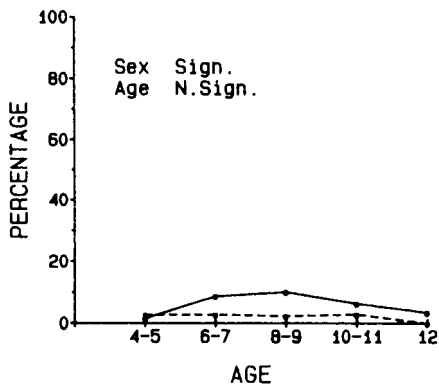
81. FEELS HURT WHEN CRITICIZED



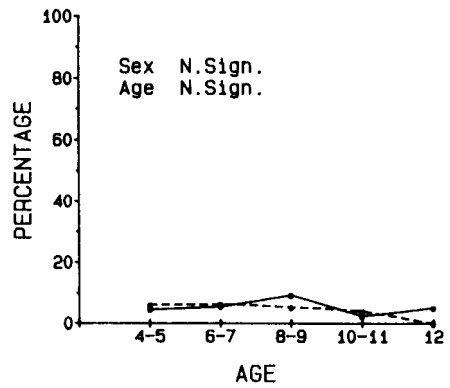
82. STEALS



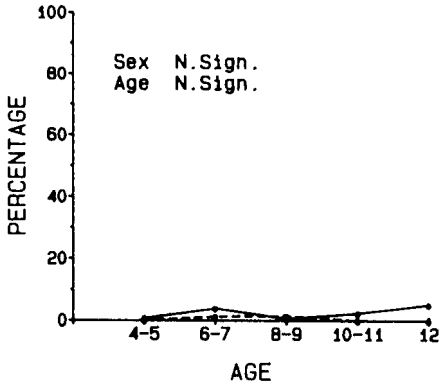
83. STORES UP UNNEEDED THINGS



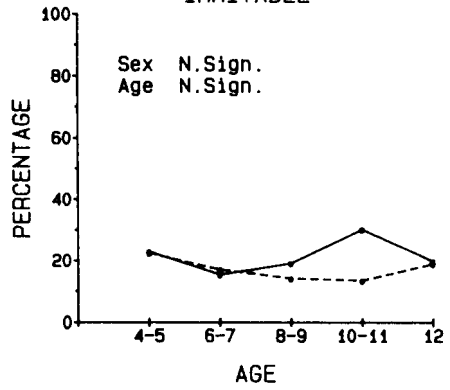
84. STRANGE BEHAVIOR



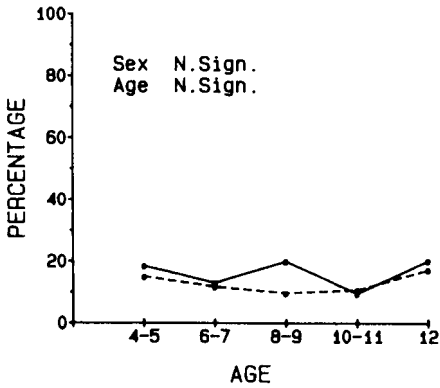
85. STRANGE IDEAS



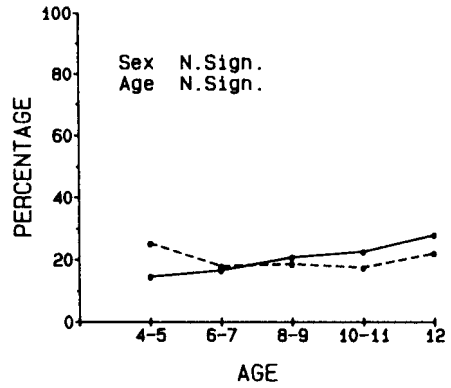
86. STUBBORN, SULLEN,
IRRITABLE



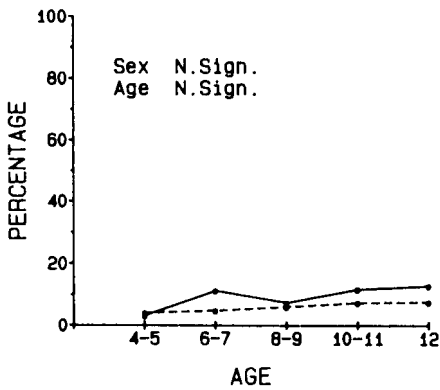
87. MOODY



88. SULKS A LOT



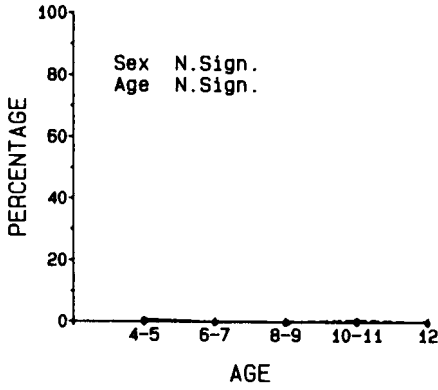
89. SUSPICIOUS



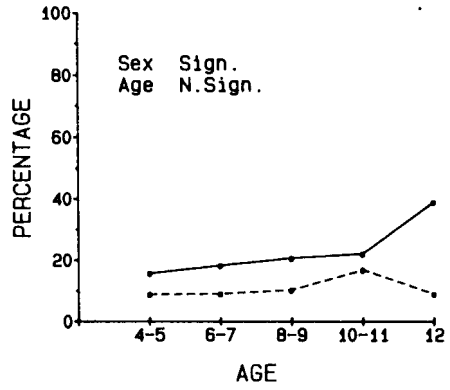
90. SWEARING



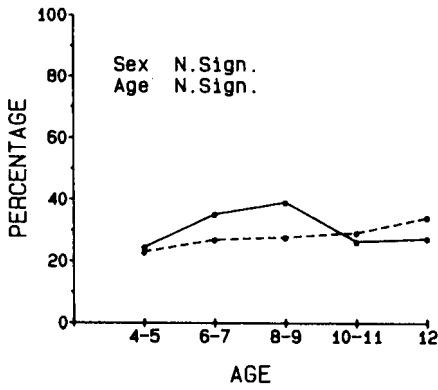
91. SUICIDAL TALK



92. UNDERACHIEVING



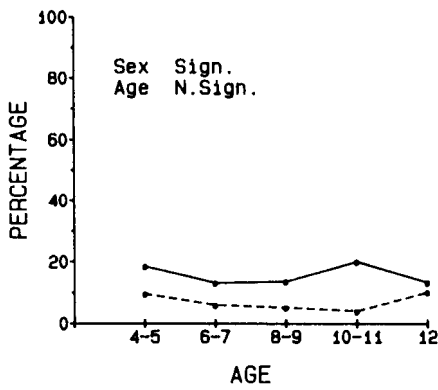
93. TALKS TOO MUCH



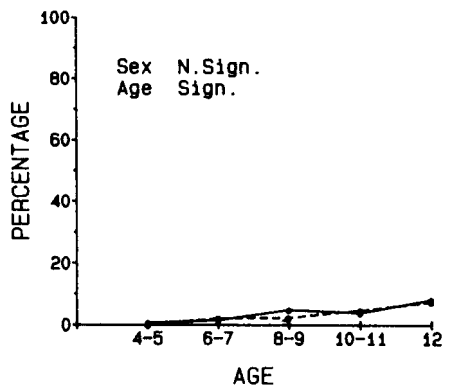
94. TEASES A LOT



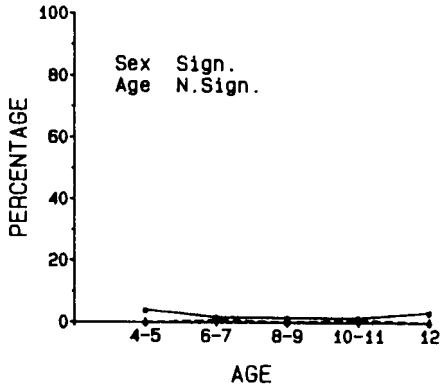
95. TEMPER TANTRUMS



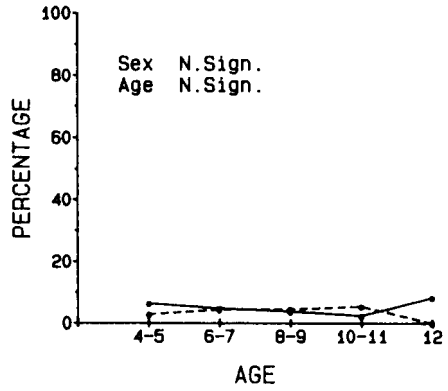
96. SEXUAL PREOCCUPATION



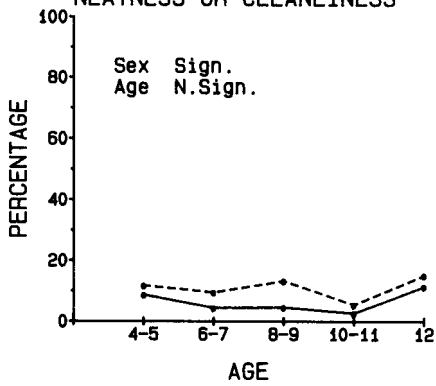
97. THREATENS PEOPLE



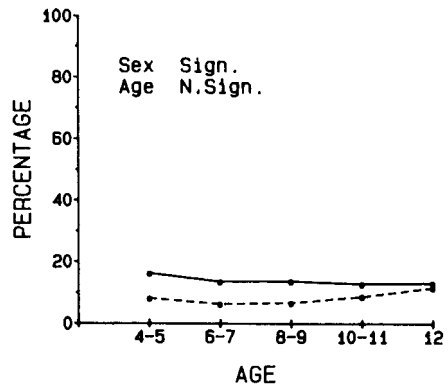
98. TARDY



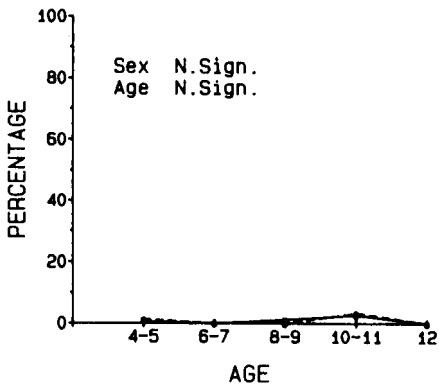
99. TOO CONCERNED WITH NEATNESS OR CLEANLINESS



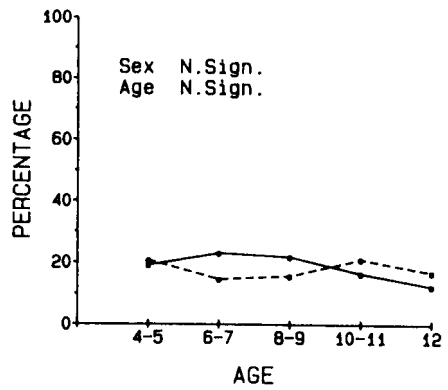
100. FAILS TO CARRY OUT TASKS



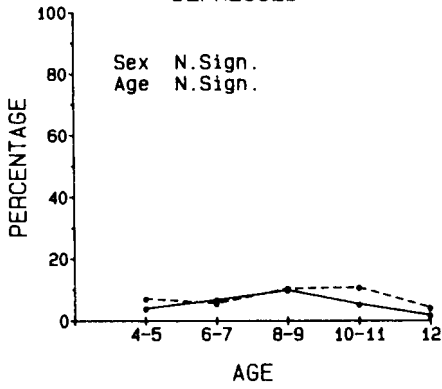
101. TRUANCY



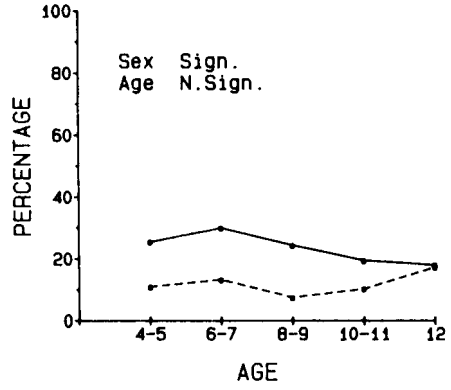
102. UNDERACTIVE



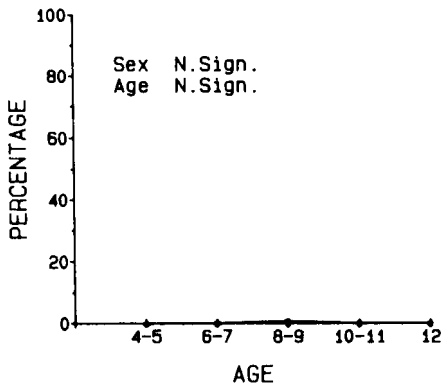
103. UNHAPPY, SAD, OR DEPRESSED



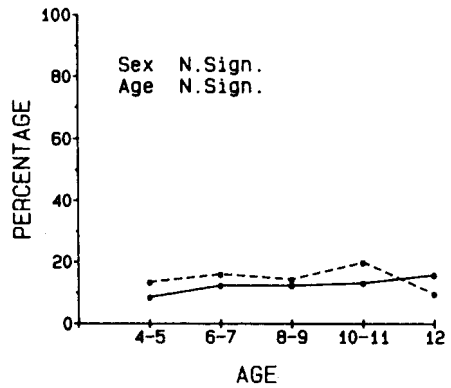
104. UNUSUALLY LOUD



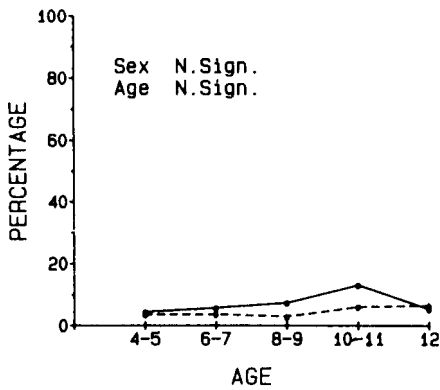
105. ALCOHOL OR DRUGS



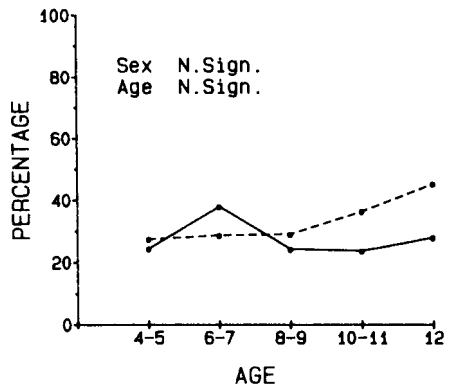
106. ANXIOUS TO PLEASE



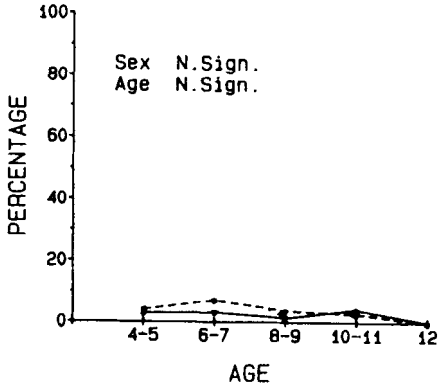
107. DISLIKES SCHOOL



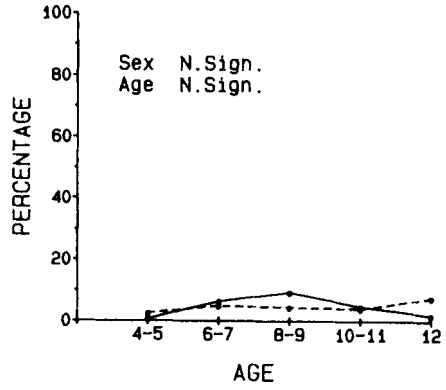
108. FEARS MISTAKES



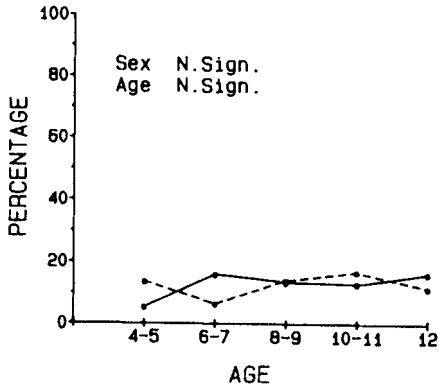
109. WHINING



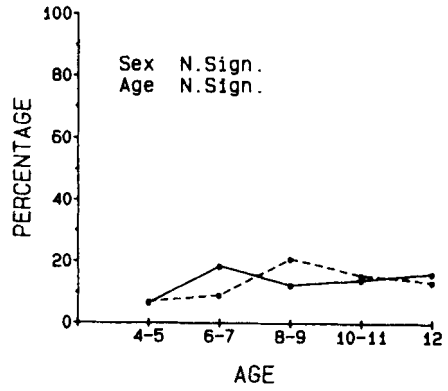
110. UNCLEAN



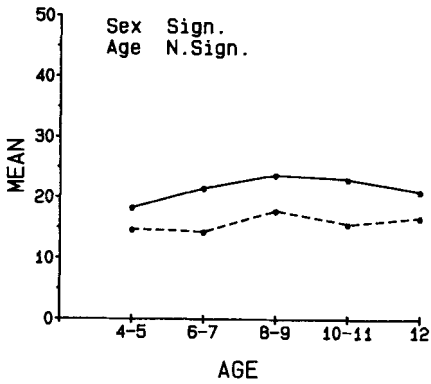
111. WITHDRAWN



112. WORRYING



113. TOTAL PROBLEM SCORE



Total problem scores

By summing all 0's, 1's and 2's for all 120 TRF items, we obtained the total problem score. Table 4 shows the means of the total scores for each age and gender for the general population sample (N = 1,162). To provide an overview, these scores are graphically portrayed for the combined two-year age groups in figure 113. The data show that boys have higher total scores than girls, whereas no consistent age effect seems to be present. Statistical analysis of age and gender effects on total problem score will be discussed together with these effects on the prevalence of specific items.

To provide norms for "healthy" children, we also report the mean total problem scores for the general population sample excluding 58 cases referred for mental health service or enrolled in special education. Researchers or clinicians who want to compare TRF scores with those of normal same-sexed age-mates can make use of this table.

Appendix B gives the cumulative frequency distribution of total problem scores for the normal sample divided by gender and age-groups 4-5 and 6-11 years. Researchers or clinicians who want to compare a particular child's scores with those of same-sexed age-mates can make use of these tables with normative data.

Gender, age and SES effects

In order to assess the effect and interaction of age and gender on the total problem score and on the scores of each problem item, we performed analyses of covariance (ANCOVA) with age and gender as main effects (a 2 (gender) x 9 (age) factorial design with SES as covariate). The results are shown in Appendix C.

With the large number of multiple tests of significance we used (namely for each of the 118 items and total problem score), a number of significant differences may arise by chance. The first measure to reduce the number of chance findings was the choice of a P value of .01 as significance level. With a sample size as large as this, even small differences may reach this significance level. We therefore determined the number of $p < .01$ findings expected by chance, using a .01 protection level. In our case five out of 119 effects could reach the .01 level of significance by chance (41). We controlled for chance findings by indicating the five significant differences having the smallest F values with a superscript in Appendix C.

According to Cohen's (42) criteria, effects accounting for 1% - 5.9% of the variance are considered small, those accounting for 5.9% - 13.8% are considered medium and effects accounting for more than 13.8% of the variance are considered large.

Table 4

TRF mean total behavior problem score by age and sex for the general population sample (N = 1162) and the normal sample (N = 1104).

| Age (group) (yrs) | General population sample | | | | | |
|----------------------|---------------------------|--------------------|-------|--------------------|------------|--------------------|
| | Boys | | Girls | | Both sexes | |
| | mean | standard deviation | mean | standard deviation | mean | standard deviation |
| 4 | 14.9 | 12.4 | 13.5 | 11.7 | 14.1 | 12.0 |
| 5 | 20.5 | 18.6 | 15.5 | 18.3 | 17.9 | 18.5 |
| 6 | 19.6 | 16.7 | 16.5 | 17.0 | 18.0 | 16.9 |
| 7 | 23.6 | 17.6 | 12.2 | 11.2 | 17.0 | 15.3 |
| 8 | 24.1 | 22.6 | 19.5 | 23.6 | 21.8 | 23.1 |
| 9 | 23.2 | 21.7 | 15.1 | 14.9 | 19.0 | 18.9 |
| 10 | 23.5 | 20.0 | 14.0 | 12.8 | 18.5 | 17.1 |
| 11 | 22.4 | 22.3 | 16.9 | 17.6 | 19.3 | 19.9 |
| 12 | 21.0 | 21.9 | 16.7 | 18.2 | 18.9 | 20.2 |

| Age (group) (yrs) | Normal sample | | | | | |
|----------------------|---------------|--------------------|-------|--------------------|------------|--------------------|
| | Boys | | Girls | | Both sexes | |
| | mean | standard deviation | mean | standard deviation | mean | standard deviation |
| 4 | 14.3 | 12.1 | 13.4 | 11.9 | 13.9 | 11.9 |
| 5 | 19.4 | 17.6 | 15.6 | 18.5 | 17.4 | 18.1 |
| 6 | 18.8 | 16.1 | 16.5 | 17.1 | 17.6 | 16.6 |
| 7 | 23.0 | 18.1 | 12.0 | 11.4 | 16.5 | 15.4 |
| 8 | 23.4 | 22.0 | 16.2 | 19.8 | 20.2 | 21.3 |
| 9 | 20.5 | 20.2 | 14.3 | 13.9 | 17.1 | 17.2 |
| 10 | 21.7 | 17.8 | 13.6 | 12.8 | 17.2 | 15.7 |
| 11 | 19.6 | 17.3 | 16.0 | 16.5 | 17.5 | 16.8 |
| 12 | 19.6 | 21.5 | 16.9 | 18.4 | 18.2 | 19.9 |

Gender differences

As can be seen in Appendix C, the effect of gender on total problem score is significant, though small. For 48 problem items, significant gender differences were found. Applying Cohen's criteria for effect size to our findings, gender differences for 31 problem items could be considered small, 4 medium and none large.

The four medium differences all showing higher scores for boys were found for the items: Bragging; Fighting; Messy work; and Showing off. The largest gender difference was found for item Bragging (12% variance accounted for).

Of the 31 problem items showing small gender differences, on only one item girls obtained higher scores than boys (Fears impulses).

For 13 items showing significant gender differences the effects accounted for < 1% of variance. After correcting for differences that may arise by chance, only 8 items (all on which boys scored higher) showed differences with a percentage of variance smaller than 1%.

Age differences

There were many fewer age differences than gender differences. No significant age effect was found for total problem score. After correction for the 5 differences that could arise by chance, 17 age effects could be considered small. On 14 items, older children scored higher, whereas nonlinear age effects were found for 3.

The number of significant interactions between age and sex (four) did not exceed chance expectation.

SES effects

Effects of socioeconomic status were partialled out by using SES as a covariate in ANCOVAs. Significant (< .01) effects were present for 18 items (13 when corrected for differences that may arise by chance) and for total problem score. On only one item (Needs to be perfect), higher SES children obtained higher scores, whereas on all other items for which SES effects were present and on total problem score, lower SES children scored higher. For 6 items and total problem score the SES effects were small according to Cohen's criteria (42), whereas for the other items showing significant SES differences the effects accounted for < 1% of variance.

Comparison between referred and nonreferred samples

The mean total problem score (49.8, sd = 29.6) for the referred sample (N = 57) was significantly higher than that for the equally sized, matched, nonreferred sample (14.9, sd = 29.6) by t-test ($t = 8.19$, $df = 12$, $p < .001$).

In order to test the discriminative power of the TRF, the sensitivity and

specificity were obtained by inspecting the cumulative frequency distributions of the total problem scores of both samples. Selecting the cutoff nearest to the 90th percentile of the normal sample, the sensitivity was 61.4% and the specificity 91.2%. The overall misclassification rate was $(38.6\% + 8.8\%)/2 = 23.7\%$.

We also compared the scores on each individual item and total problem-score obtained by children in the referred sample with those obtained by children in the nonreferred sample by computing ANCOVAs using a 2 (referred vs nonreferred) x 8 (age) x 2 (gender) factorial design and SES as covariate.

On 79 of the 118 items referred children scored higher than the nonreferred children at a $p < .05$ level. For 22 items the effect of referral status was large according to Cohen's criteria (accounting for more than 13.8% of the variance). The effect of referral status was largest for the following items: Demands attention; Impulsive; Poor school work and Stares blankly.

For the total problem score referral status accounted for 37% ($p < .001$) of the variance.

DISCUSSION

Prevalence of specific items

When we evaluate the prevalence rates for individual problems as reported by teachers, the following issues may be important.

1. Specific behavioral-emotional problems need not necessarily reflect psychiatric disorder; the rarity of a particular problem may be related however, to the severity of the problem (e.g. suicidal talk, stealing, etc.).
2. As yet, no single approach is totally superior to all others for assessing children's behavioral-emotional problems; every procedure inevitably involves subjective judgments and particular interactions with the child.
3. Not one informant can provide data on all aspects of children's functioning; different sources of information can validly contribute to the picture of a child's functioning.
4. There is no diagnostic system that fully satisfies reliability and validity criteria; in fact data on individual symptoms may help to improve diagnostic systems through the assessment of covariation among individual symptoms in order to define syndromes.
5. Differences in item wording may cause differences in rates across different studies; data need to be obtained through the same assessment procedures to detect general phenomena.

Taking account of these issues, the prevalence rates depicted in figures 1-113, provide norms for a wide variety of clinically important aspects of children's functioning at school.

Comparison with other studies

Due to different item wording, age range or scoring format we were able to compare our item prevalence rates with only five of the eight studies listed in table 1 for which item frequencies were reported. From the studies by Roberts et al. (26), Rutter et al. (31), Schultz et al. (35), Sheperd et al. (34) and Werry et al. (37), we were able to select 30 items from one or more studies that were reasonably comparable to ours. Of the 134 comparisons made between our data and those from the other studies, 26 (19%) showed differences in prevalence rates of at least 10%. For 16 of these differences our rates were higher and for 10 the rates of our study were lower (see table 5). This difference was not significant ($\chi^2 = .69$, $df = 1$, ns). Dutch children were scored higher than children in at least two other studies on: Can't concentrate and Hyperactive. These two items were also among the items on which Dutch parents scored their children higher than parents from other studies (2,5). On two items, Feels worthless and Behaves irresponsibly, Dutch children were scored lower than children in at least two other studies.

Whether these differences reflect true differences in prevalence rates or

differences in teachers' and parents' thresholds for reporting these problems cannot be concluded from our data. The main conclusion we can draw from this comparison is that our data did not differ to a large or consistent extent from other studies' data that were comparable to ours.

Table 5

Comparison with other studies. Differences in prevalence rates of 10% or more.

| Study | Nr. of comparisons | Dutch rates higher | | Other study's rates higher | |
|------------------------|--------------------|-------------------------------------|---------------------------------------|---|----------------------------------|
| | | Boys | Girls | Boys | Girls |
| Roberts et al. (26) | 10 | Can't conc. Fighting | Can't conc. | | |
| Rutter et al. (31) | 32 | Can't conc. Hyperactive Cruel | Can't conc. Hyperactive Fearful | Worrying | |
| Schultz et al. (35) | 38 | Can't conc. Secretive | Secretive | Feels worthless Irresponsible | Feels worthless Irresponsible |
| Sheperd et al. (34) | 16 | Hyperactive | Hyperactive | | |
| Werry et al. (37) | 38 | Secretive | Secretive | Disobedient Feels worthless Irresponsible | Feels worthless Irresponsible |

Comparison with Achenbach and Edelbrock's data

Because we used the same instrument and the same general methodology as Achenbach and Edelbrock (13), comparison between the results of both studies can be made more precisely than the comparisons with other studies described in the previous section. The mean total problem scores of 17.6 for Dutch and 19.3 for American 6-11-year-old nonreferred children, showed no significant difference ($F < 1$) (43). Table 6 shows the mean problem scores for both nationalities and different age groups.

Table 6

Mean total problem scores of nonreferred Dutch and American children.

| Age | Sex | Dutch | | American | |
|-------|-------|-------|------|----------|------|
| | | N | Mean | N | Mean |
| 6-7 | Boys | 110 | 20.7 | 98 | 22.8 |
| | Girls | 143 | 14.2 | 99 | 16.6 |
| 8-9 | Boys | 127 | 22.2 | 99 | 24.0 |
| | Girls | 123 | 15.2 | 97 | 16.7 |
| 10-11 | Boys | 114 | 20.6 | 96 | 19.7 |
| | Girls | 144 | 14.9 | 97 | 15.6 |
| 6-11 | Boys | 351 | 21.2 | 293 | 22.2 |
| | Girls | 410 | 14.8 | 293 | 16.3 |

On the level of individual problem items, we will report here on differences between both studies' prevalence rates of 10% or more. A more rigorous statistical analysis of these differences has been reported by Achenbach et al. (43).

The frequency with which each problem was reported in the normal (non-referred) samples in both studies was compared for each gender and age intervals 6-7, 8-9 and 10-11. A total of 120 (items) x 3(age) x 2(gender) = 720 comparisons could be made. Scores of 1 and 2 were combined. For 124 (17%) comparisons, differences in prevalence rates of 10% or more were found. This is somewhat less than the 19% comparisons for which differences were found between Dutch and American prevalence rates of parent reported problem behavior in children aged 4-16 (2).

Although total problem scores between both nationalities did not show significant differences, the proportion of items on which Dutch or American children scored higher differed ($\chi^2 = 3.90$, $df = 1$, $p < .05$). For 73 comparisons the American rates were higher and for 51 the Dutch rates were. In order to investigate possible consistencies in these differences associated with type of problem, we looked at their distribution across empirically derived broad band syndromes. Because our TRF items were similar to the original ones used by Achenbach and Edelbrock (13), we could make use of the results of their factor analyses. The authors factor analyzed TRF's filled out by teachers of large samples of clinically referred children. For each gender and age groups 6-11 and 12-16 years, narrow band as well as broad band factors were obtained. The broad band syndrome called externalizing involves problem behavior directed towards the environment, whereas the internalizing syndrome

encompasses symptoms mainly involving internal distress and conflicts.

We assessed the association of items showing nationality differences with the externalizing or internalizing syndromes. Of the 54 items showing nationality differences in prevalence rates of 10% or more that could be classified, no significant difference between proportion of internalizing versus externalizing could be found ($\chi^2 = .38, df = 1, ns$).

Those items showing the largest nationality differences are listed in table 7. As a criterion we chose to report those items on which nationality differences occurred for at least 3 of the 6 different gender/age groups.

Table 7

Items on which American and Dutch studies disagreed 10% or more in at least three different age/gender groups.

| American rates higher | | Dutch rates higher |
|---------------------------------|----------------------------|--------------------|
| Acts too young | Poor schoolwork | Bragging |
| Hums or other odd noises | Disrupts class discipline | Likes to be alone |
| Fails to finish things | Self-consciousness | Nervous |
| Fidgets | Feels hurt when criticized | Secretive |
| Difficulty following directions | Underachieving | Sulks a lot |
| Disturbs other pupils | Fails to carry out tasks | |
| Fears impulses | | |

The main conclusion we can draw from the comparisons between our study and others is that the similarities by far outweigh the differences, suggesting that teacher reported behavioral-emotional problems of children tend to be rather uniformly distributed across the various countries. However, all countries for which comparisons were carried out have a western culture and, except The Netherlands, are all English speaking. Comparisons with teacher ratings of children from countries having different socio-cultural backgrounds may reveal possible environmental influences.

Gender differences

Teachers reported significantly more behavioral-emotional problems for boys than for girls. Although the effect size is small according to Cohen's criteria, the mean total problem scores are higher for boys than for girls.

On 48 individual items (43 when corrected for differences that may arise by chance), gender differences were found. Boys scored higher on 45 items, girls on 3 items (male/female ratio of 15). To determine consistencies of type of

problems scored higher for one gender than the other, we examined the association of these problems with externalizing versus internalizing syndromes. Table 8 shows the results. As can be seen the majority of items on which boys score higher were associated with the externalizing syndrome ($\chi^2 = 6.71$, $df = 1$, $p < .01$). The number of items on which girls score higher was too small to detect consistencies in problem type.

Many of the externalizing items on which boys scored higher than girls involved disruptive and aggressive behaviors. Boys also score higher on items related to academic problems such as: Difficulty learning; Poor school work and Underachieving.

If we compare our results with those from Achenbach and Edelbrock (13), for 6-16-year-old children, we found gender effects on total problem score in the same direction and of the same size across both studies. In the American study boys were also found to show more generally disturbing behavior and poorer school functioning than girls.

Next we compared gender differences reported by teachers with those reported by parents.

In our earlier study of behavioral-emotional problems reported by parents of 4-16-year-old children (2), parents also reported more problems for boys than girls, but to a lesser extent than teachers. Gender accounted for <1% of the variance in total problem scores reported by parents, compared to 3% of variance in total problem scores reported by teachers. On the level of individual items, parents rated boys higher on 30 CBCL-items and girls on 11. This male/female ratio of nearly 3 to 1 is much smaller than the ratio of 15 to 1 found for the teacher ratings.

In order to assess consistencies in gender effects across teacher and parent ratings, we took the mean of the percentage of variance accounted for by gender effects in both studies for each of the 94 corresponding items on the TRF and CBCL respectively. It should be noted that consistencies of gender effects across teacher versus parent ratings do not necessarily imply consistencies in individual children's behavior. They merely indicate general trends in gender differences. Table 9 shows the 13 items for which the mean effect size for gender differences across teacher and parent reported problems was at least 1%. According to Cohen's criteria only Bragging and Fighting can be considered medium effects. The only item on which girls were scored significantly higher by parents and teachers was Too concerned with neatness or cleanliness. However, this effect accounted for less than 1% of the variance. Although parents also scored boys higher than girls especially on externalizing items, teacher scores revealed a larger gender difference on externalizing versus internalizing items.

Table 8

Problem items with gender differences ($p < .01$) in ANCOVAs.
 Association with Externalizing and Internalizing syndromes*.

| | Boys scoring higher than Girls | | Neither |
|--------------------------|--------------------------------|----------------------------|-----------------|
| | Externalizing | Internalizing | |
| Acts too young | | Feels hurt when criticized | Accident prone |
| Hums or other odd noises | Difficulty learning | | Sleeps in class |
| Argues | Talks out of turn | | Speech problem |
| Fails to finish things | Attacks | | |
| Defiant | Apathetic, unmotivated | | |
| Bragging | Poor schoolwork | | |
| Can't concentrate | Clumsy | | |
| Hyperactive | Disruptive | | |
| Fidgets | Screams | | |
| Cruelty | Messy work | | |
| Destroys own things | Showing off | | |
| Can't follow directions | Explosive, unpredictable | | |
| Disobedient | Inattentive | | |
| Disturbs others | Hoarding | | |
| Lacks guilt | Swearing | | |
| Fighting | Underachieving | | |
| Impulsive | Teases | | |
| Lying, cheating | Temper tantrums | | |
| Nervous | Threatens | | |
| Nervous movements | Doesn't carry out tasks | | |
| | Loud | | |
| | Girls scoring higher than Boys | | |
| | Externalizing | Internalizing | |
| | Jealous | Fears impulses | |
| | | Too neat | |

*Empirically derived broad band syndromes; source Achenbach and Edelbrock (13).
 Items assigned on the basis of their highest factor loading.

Table 9

Items showing the largest gender differences* in ANCOVAs of teacher and parent reported problems. The mean percentage of variance accounted for is shown in brackets (v).

| Item | (v) | Item | (v) |
|-------------|------|---------------------|-----|
| Bragging | (12) | Temper tantrums | (2) |
| Fights | (6) | Can't concentrate | (2) |
| Showing off | (5) | Swearing | (2) |
| Hyperactive | (3) | Destroys own things | (1) |
| Teases | (3) | Accident prone | (1) |
| Loud | (3) | Speech problems | (1) |
| Clumsy | (2) | | |

*On all items boys scored higher than girls

The general tendency of boys to show somewhat more aggressive and generally disturbing behavior than girls may be more salient in the school environment with its more task oriented and structured atmosphere than at home or outdoors. Teachers may especially focus on behaviors interfering with classroom functioning and less on behavior that may cause distress to the child but do not clearly interfere with the child's academic functioning.

In conclusion, it was found that teachers, and to a lesser extent parents, reported more problem behavior for boys than for girls, the majority of the problems being associated with generally disturbing behavior and poor academic functioning. These gender differences in adult reported behavioral-emotional problems may be associated with the higher mental health referral rates for boys than for girls during the elementary school period.

Age differences

Age showed no significant effect on total problem scores across ages 4-12. The number of items for which significant ($p < .01$) age differences were found was much smaller than that for which gender differences were found. Older children scored higher on 15 items, whereas younger children scored higher on only two items. After correction for findings that may arise by chance, the number of items on which older children scored higher was 14, whereas younger children scored higher on none. According to Cohen's criteria, all of the effects could be considered small. The largest effect (4% variance accounted

for) was found for Difficulty learning.

Table 10 shows the distribution of items for which linear age effects were found, across externalizing versus internalizing syndromes. Although more items seem to be associated with the externalizing than with the internalizing syndrome, this difference was not significant ($\chi^2 = 2.27, df = 1, ns$). If we take a closer look at the individual items, we find that older children tended to score higher on items associated with academic functioning, somatic functioning, social functioning and inner feelings. Feels unloved was reported for 4% of 6-7-year-olds and for 11% of 10-11-year-olds. Feels worthless was reported for 6% of the 6-7-year-olds and for 13% of the 10-11-year-olds. The increase with age in the occurrence of these emotions (small effects according to Cohen's criteria) may reflect children's socio-emotional development in which differentiation of interpersonal relationships is accompanied by differentiation of self-concepts. As children grow older, they show increasing disparity between real-self and ideal-self concepts (44). Achenbach and Edelbrock (13) found a small age effect in the same direction for the item Feels worthless, but not for Feels unloved. In general, however, we cannot draw firm conclusions concerning developmental aspects, because in both our study and Achenbach and Edelbrock's study, older children were found to score higher on only four items. These items are: Feels worthless; Not liked; Poor schoolwork and Sexual preoccupation. Age differences showed much less consistency across both nationalities than sex differences, although both studies found that age differences were smaller and less numerous than sex differences. The reverse was true for parent reported problems in which both Dutch (2) and American (3) studies reported somewhat larger age differences than sex differences. Parents reported more problems for younger than for older children. The finding that teachers' reports show greater stability across ages than parents' reports may reflect a tendency of teachers to assess children's behavior in relation to norms implicitly adjusted to particular age levels, whereas parents may assess their child's behavior in more absolute terms.

There were only three items on which teachers as well as parents scored older children significantly higher than younger children. These items were: Feels worthless, Headaches and Poor school work. No other age effects showed consistencies.

The increase in teacher scores across ages 6-11 on the items Poor school work and Difficulty learning can be attributed to the increase in scores for girls. On both items boys showed small, non-linear age differences, whereas for girls the rates doubled for both items. At 10-11 years boys and girls were rated nearly equal.

In conclusion, it was found that teacher reports on behavioral-emotional problems were less sensitive to developmental differences than parent reports. It is likely that teachers use their own age norms against which a child's behavior is assessed.

SES differences

Teachers reported more problems for lower SES children than for higher SES children. Higher SES children scored higher only on the item Needs to be perfect, whereas lower SES children scored significantly higher on 18 items (13 when corrected for chance effects) and on total problem score. Most other studies reporting SES effects on the prevalence of teacher reported problems also found higher rates for lower SES children. Lower SES children were found to have poorer school achievement (27) and more language delays (45). Our data also showed that problems in academic functioning were more prevalent in lower SES children. Significantly higher scores for lower SES children on the items Difficulty learning and Poor school work were also found in Achenbach and Edelbrock's (13) study.

Table 11 shows the association with externalizing and internalizing syndromes of those items on which lower SES children scored higher. As can be seen, more items are associated with the externalizing than with the internalizing syndrome ($\chi^2 = 8.33$, $df = 1$, $p < .01$). In a number of areas such as attention, control of impulses and academic -, social -, and motor functioning, lower SES children showed less organized behavior than higher SES children.

Lower SES children in our sample were also found to show more behavioral-emotional problems as reported by their parents (2). On the following items, lower SES children obtained higher scores in both parents' and teachers' reports: Confused; Disobedient; Is teased; Swearing; Teases; Sex preoccupation; Too dependent; Whining and Picking. Lower SES parents reported their children to be less competent in areas of social adjustment and school achievement.

In conclusion, lower SES children showed more behavior problems at home and at school and performed less well in school. The factors involved in this association are poorly understood. Higher rates of stressful life events, unfavourable housing (31,45), less adequate parenting (46) and language delay (45) are examples of often intercorrelated factors that may affect childrens' psycho-educational functioning.

Differences related to academic functioning

Of the total sample of 1,162 children, 39 (3.4%) attended schools for special education. For the 6-12-year-olds this percentage is 4.3, which is somewhat lower than the 6% national estimate for this age group (39). More boys (66%) than girls received special education ($\chi^2 = 4.12$, $df = 1$, $p < .05$).

The mean total problem score of 41.9 ($sd = 28.7$) for children receiving special education was more than twice as high as the mean total problem score of 18.1 ($sd = 17.8$) for the nonreferred sample of 6-11-year-olds. This difference was significant ($t = 4.49$, $df = 798$, $p < .001$), indicating that the TRF differentiates significantly between children in regular schools and those in special schools, despite the inclination of teachers to adopt individual

Table 10

Problem items with age differences ($p < .01$) in ANCOVAs.
Association with Externalizing and Internalizing syndromes [★].

| Younger children scoring higher | | |
|---------------------------------|----------------------------|----------------------------|
| Externalizing | Neither | |
| Fighting | Prefers older children | |
| Older children scoring higher | | |
| Externalizing | Internalizing | Neither |
| Confused | Feels unloved ^B | Feels unloved ^G |
| Disturbs others | Feels worthless | Is teased ^B |
| Feels persecuted | Secretive | Not liked ^G |
| Is teased ^G | | Overweight |
| Not liked ^B | | Headaches |
| Difficulty learning | | Stomachaches |
| Poor schoolwork | | Sexual preoccupation |
| Messy work | | |

[★]Empirically derived broad band syndromes; source Achenbach and Edelbrock (13). Items assigned on the basis of highest factor loading.
B = Boys; G = Girls.

Table 11

Association with Externalizing and Internalizing syndromes of behavior problem items on which lower[★] SES children score higher ($p < .01$) in ANCOVAs^{★★}.

| Externalizing | Internalizing | Neither |
|------------------------|---------------|------------------------|
| Confused | Too dependent | Accident prone |
| Disobedient | | Is teased ^B |
| Disturbs others | | Picking |
| Is teased ^G | | Clumsy ^G |
| Impulsive | | Sexual preoccupation |
| Difficulty learning | | Whining |
| Poor schoolwork | | Unclean |
| Clumsy ^B | | |
| Inattentive | | |
| Swearing | | |
| Teases others | | |

[★]On only one (internalizing) item (Needs to be perfect) higher SES children scored higher.

^{★★}Empirically derived broad band syndromes; source Achenbach and Edelbrock (13). Items assigned on the basis of highest factor loading.
B = Boys; G = Girls.

norms for their own classroom situation.

We were also interested in the relationship between TRF scores and cognitive functioning. We were able to assess this relation by using data from our previously reported prevalence study (33) in which 116 children selected from the general population on the basis of their CBCL or TRF scores were interviewed. For 110 children, we also obtained IQ scores by administering the WISC-R short form (47). A moderate, though significant, negative correlation of $-.34$ ($p < .001$) was found between IQ and TRF scores. The negative correlation between IQ and CBCL scores was not significant ($r = -.17$).

The TRF contains a number of items on problems in academic functioning (Difficulty following directions; Difficulty learning; Underachieving) that replace CBCL items of a more general nature or confined to the house situation (Disobedient at home; Constipated; Sleepwalking). The TRF items capturing problems in academic functioning all load on the same factor (Inattentive) in factor analyses of boys and girls 6-11-years (13), indicating their interrelatedness.

The TRF is aimed more at capturing problems in academic functioning than is the CBCL, because teachers are in a better position than parents to observe these problems. On the other hand, it may be that less intelligent children show more deviant behavior in task oriented situations.

The findings stress the position outlined in the introduction that different informants such as teachers and parents may provide different information that is relevant to particular situations or to specific aspects of the child's functioning.

The discriminative power of the TRF

Both from a practical-clinical and research point of view it may be important to obtain standardized information from teachers for the following reasons:

1. Standardized information can be obtained with a minimum of time involvement; direct contact with the teacher may follow initial assessment, if necessary;
2. Standardized information can be easily stored in files or computers and is readily surveyable by clinicians or researchers;
3. Standardized information can be compared with information on other cases;
4. Standardized information from one source (e.g. teachers) can be compared with that from another (e.g. parents) and
5. Standardized information from one mental health or research setting can be compared with that from others.

In our review of the five most widely used standardized teacher rating forms for which psychometric data were available, we concluded that the TRF has the following advantages not shared by others.

1. The TRF covers a broad range of behavioral-emotional problems;
2. The TRF is normed for different age and sex groups;
3. The TRF is scorable on a profile portraying an individual child's

relative position compared to normal age- and sex-mates on a number of empirically derived syndromes and

4. The TRF has the same format as the analogous parent questionnaire, the CBCL (Child Behavior Checklist).

An important validity criterion of an instrument like the TRF which was designed to measure degree and type of child psychopathology, is its potential to discriminate between psychiatrically disturbed and non-disturbed children.

In the previous paragraph we reported that the TRF differentiates between children who attend special schools and those who go to regular schools. To evaluate the contribution of our findings, we must also consider the findings of our previously reported studies (2,33). Using clinical judgment as a morbidity criterion, the CBCL and TRF were found to be nearly equal in correctly identifying disturbed and non-disturbed children. The percentage correct classification was 73 for the CBCL and 72 for the TRF when only the problem items were used.

We also found that the CBCL discriminated well between large samples of referred and nonreferred children (2). At this moment we do not have TRF data on large samples of clinically referred Dutch children. Nevertheless, preliminary findings on a small (N = 57) sample of referred children matched for age, gender and SES with a sample of nonreferred children (N = 57) supported the discriminative validity of the TRF, as can be concluded from the percentage correct classification of 76% using a cutoff point for total problem scores corresponding with the 90th percentile of nonreferred children. Referral status showed a significant effect in ANCOVAs of the referred and nonreferred samples for 79 of the 118 items and for total problem score.

The data from our previous and present studies support the discriminative validity of the TRF.

SUMMARY

Standardized teacher reports on children's behavioral-emotional problems can provide information on areas of children's functioning not readily accessible to other informants such as parents or clinicians. Once we accept the generally low agreement between different informants as inevitable, we may make use of the different types of information available from different sources. To do so we need base-line data from different informants that take account of key demographic variables such as gender, age and socio-economic status.

In a previous report we provided such data for parent reported behavioral-emotional problems in children aged 4 through 16. In the present study we (1) reported on the prevalence of a wide range of specific behavioral-emotional problems reported by teachers in a representative sample of 1,162 Dutch children aged 4 through 12; (2) identified differences related to demographic variables and (3) compared our data with those from other population based surveys.

We used the Achenbach Teacher's Report Form (TRF) to collect our data because a comparison between different standardized teacher assessment instruments showed the TRF to be the most promising due to its solid psychometric background and its ready applicability. Our data supported the TRF's reliability and validity.

For each of the 118 behavioral-emotional problems, the prevalence rates were presented graphically for both genders in two-year age groups. ANCOVAs were performed to assess the main effects and interaction of age and gender with SES as covariate for each problem item and total problem score.

Of the demographic variables gender showed most numerous effects on problem items and showed the largest effect on total problem score. The main findings were:

1. Boys obtained higher scores than girls.
2. Many items on which boys scored higher are related to socially disapproved behavior.
3. Boys obtained higher scores on concentration, attention and hyperactivity problems.
4. Boys (especially the younger ones) scored higher on problems related to academic functioning.
5. Teacher reports revealed larger gender differences than parent reports.
6. Both parents and teachers scored boys higher on items predominantly associated with the externalizing syndrome.
7. More boys than girls were attending education in special schools.

Age showed no significant effect on total problem score, although older children were scored higher on a number of items associated with academic, somatic and social functioning and with emotions related to inner feelings.

The much fewer age differences reported by teachers compared with parents may indicate that teachers are less sensitive than parents to developmental differences due to a tendency to set norms for the particular grade level they

teach.

Lower SES children were reported to show more problems than higher SES children. Teachers scored lower SES children especially higher on externalizing problems that reflected problems in academic and social functioning and behaviors that are disturbing in the classroom setting. As parent reports also showed somewhat more behavioral problems and fewer social competencies in lower SES children, it was concluded that lower SES children are disadvantaged across a number of important areas of functioning.

Teacher's ratings of behavioral-emotional problems were significantly related to level of cognitive abilities and to referral for special education.

Although comparison with other population based studies on the prevalence of teacher reported problems revealed a number of differences for problem rates, the overall conclusion was that our data did not differ systematically from others. The findings support the applicability of the results of the TRF across both the U.S. and The Netherlands.

APPENDIX A

— for office use only —
IDENTIFICATION # _____

CHILD BEHAVIOR CHECKLIST - TEACHER'S REPORT FORM

| | | | |
|-------------|--|----------------------------|--------------|
| PUPIL'S AGE | PUPIL'S SEX <input type="checkbox"/> Boy <input type="checkbox"/> Girl | ETHNIC GROUP OR RACE | PUPIL'S NAME |
| GRADE | THIS FORM FILLED OUT BY <input type="checkbox"/> Teacher (name) _____ <input type="checkbox"/> Counselor (name) _____ <input type="checkbox"/> Other (specify) _____ name: | | SCHOOL |
| DATE | | | |

PARENTS' TYPE OF WORK (Please be specific — for example, auto mechanic, high school teacher, homemaker, laborer, lathe operator, shoe salesman, army sergeant.)

FATHER'S TYPE OF WORK _____ MOTHER'S TYPE OF WORK _____

I. How long have you known this pupil?

II. How well do you know him/her? Very Well Moderately Well Not Well

III. How much time does he/she spend in your class per week?

IV. What kind of class is it? (Please be specific, e.g., regular 5th grade, 7th grade math, etc.)

V. Has he/she ever been referred for special class placement, services, or tutoring?

No Don't Know Yes — what kind and when?

VI. Has he/she ever repeated a grade?

No Don't Know Yes — grade and reason

VII. Current school performance — list academic subjects and check appropriate column:

| Academic subject | 1. Far below grade | 2. Somewhat below grade | 3. At grade level | 4. Somewhat above grade | 5. Far above grade |
|------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 1. _____ | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 2. _____ | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 3. _____ | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 4. _____ | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 5. _____ | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| 6. _____ | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

*Copyright 1980 Thomas M. Achenbach and Craig Edelbrock
Thomas M. Achenbach, Ph.D.
Center for Children, Youth, & Families
University of Vermont
1 South Prospect St.
Burlington, VT 05401

985 Edition

Below is a list of items that describe pupils. For each item that describes the pupil now or within the past 2 months, please circle the 2 if the item is **very true** or **often true** of the pupil. Circle the 1 if the item is **somewhat** or **sometimes true** of the pupil. If the item is **not true** of the pupil, circle the 0. Please answer all items as well as you can, even if some do not seem to apply to this pupil.

| 0 = Not True (as far as you know) | | | 1 = Somewhat or Sometimes True | | | 2 = Very True or Often True | | |
|-----------------------------------|---|---|--|---|---|-----------------------------|--|--|
| 0 | 1 | 2 | 1. Acts too young for his/her age | 0 | 1 | 2 | 31. Fears he/she might think or do something bad | |
| 0 | 1 | 2 | 2. Hums or makes other odd noises in class | 0 | 1 | 2 | 32. Feels he/she has to be perfect | |
| 0 | 1 | 2 | 3. Argues a lot | 0 | 1 | 2 | 33. Feels or complains that no one loves him/her | |
| 0 | 1 | 2 | 4. Fails to finish things he/she starts | 0 | 1 | 2 | 34. Feels others are out to get him/her | |
| 0 | 1 | 2 | 5. Behaves like opposite sex | 0 | 1 | 2 | 35. Feels worthless or inferior | |
| 0 | 1 | 2 | 6. Defiant, talks back to staff | 0 | 1 | 2 | 36. Gets hurt a lot, accident-prone | |
| 0 | 1 | 2 | 7. Bragging, boasting | 0 | 1 | 2 | 37. Gets in many fights | |
| 0 | 1 | 2 | 8. Can't concentrate, can't pay attention for long | 0 | 1 | 2 | 38. Gets teased a lot | |
| 0 | 1 | 2 | 9. Can't get his/her mind off certain thoughts; obsessions (describe): _____ | 0 | 1 | 2 | 39. Hangs around with others who get in trouble | |
| 0 | 1 | 2 | 10. Can't sit still, restless, or hyperactive | 0 | 1 | 2 | 40. Hears things that aren't there (describe): _____ | |
| 0 | 1 | 2 | 11. Clings to adults or too dependent | 0 | 1 | 2 | 41. Impulsive or acts without thinking | |
| 0 | 1 | 2 | 12. Complains of loneliness | 0 | 1 | 2 | 42. Likes to be alone | |
| 0 | 1 | 2 | 13. Confused or seems to be in a fog | 0 | 1 | 2 | 43. Lying or cheating | |
| 0 | 1 | 2 | 14. Cries a lot | 0 | 1 | 2 | 44. Bites fingernails | |
| 0 | 1 | 2 | 15. Fidgets | 0 | 1 | 2 | 45. Nervous, high-strung, or tense | |
| 0 | 1 | 2 | 16. Cruelty, bullying, or meanness to others | 0 | 1 | 2 | 46. Nervous movements or twitching (describe): _____ | |
| 0 | 1 | 2 | 17. Daydreams or gets lost in his/her thoughts | 0 | 1 | 2 | 47. Overconforms to rules | |
| 0 | 1 | 2 | 18. Deliberately harms self or attempts suicide | 0 | 1 | 2 | 48. Not liked by other pupils | |
| 0 | 1 | 2 | 19. Demands a lot of attention | 0 | 1 | 2 | 49. Has difficulty learning | |
| 0 | 1 | 2 | 20. Destroys his/her own things | 0 | 1 | 2 | 50. Too fearful or anxious | |
| 0 | 1 | 2 | 21. Destroys property belonging to others | 0 | 1 | 2 | 51. Feels dizzy | |
| 0 | 1 | 2 | 22. Difficulty following directions | 0 | 1 | 2 | 52. Feels too guilty | |
| 0 | 1 | 2 | 23. Disobedient at school | 0 | 1 | 2 | 53. Talks out of turn | |
| 0 | 1 | 2 | 24. Disturbs other pupils | 0 | 1 | 2 | 54. Overtired | |
| 0 | 1 | 2 | 25. Doesn't get along with other pupils | 0 | 1 | 2 | 55. Overweight | |
| 0 | 1 | 2 | 26. Doesn't seem to feel guilty after misbehaving | 0 | 1 | 2 | 56. Physical problems without known medical cause: | |
| 0 | 1 | 2 | 27. Easily jealous | 0 | 1 | 2 | a. Aches or pains | |
| 0 | 1 | 2 | 28. Eats or drinks things that are not food (describe): _____ | 0 | 1 | 2 | b. Headaches | |
| | | | _____ | 0 | 1 | 2 | c. Nausea, feels sick | |
| | | | _____ | 0 | 1 | 2 | d. Problems with eyes (describe): _____ | |
| 0 | 1 | 2 | 29. Fears certain animals, situations, or places other than school (describe): _____ | 0 | 1 | 2 | e. Rashes or other skin problems | |
| | | | _____ | 0 | 1 | 2 | f. Stomachaches or cramps | |
| | | | _____ | 0 | 1 | 2 | g. Vomiting, throwing up | |
| 0 | 1 | 2 | 30. Fears going to school | 0 | 1 | 2 | h. Other (describe): _____ | |
| | | | | | | | _____ | |

| 0 = Not True | | | 1 = Somewhat or Sometimes True | | | 2 = Very True or Often True | | |
|--------------|---|---|---|---|---|-----------------------------|---|--|
| 0 | 1 | 2 | 57. Physically attacks people | 0 | 1 | 2 | 84. Strange behavior (describe): _____ | |
| 0 | 1 | 2 | 58. Picks nose, skin, or other parts of body (describe): _____ | 0 | 1 | 2 | 85. Strange ideas (describe): _____ | |
| 0 | 1 | 2 | 59. Sleeps in class | 0 | 1 | 2 | 86. Stubborn, sullen, or irritable | |
| 0 | 1 | 2 | 60. Apathetic or unmotivated | 0 | 1 | 2 | 87. Sudden changes in mood or feelings | |
| 0 | 1 | 2 | 61. Poor school work | 0 | 1 | 2 | 88. Sulks a lot | |
| 0 | 1 | 2 | 62. Poorly coordinated or clumsy | 0 | 1 | 2 | 89. Suspicious | |
| 0 | 1 | 2 | 63. Prefers being with older children | 0 | 1 | 2 | 90. Swearing or obscene language | |
| 0 | 1 | 2 | 64. Prefers being with younger children | 0 | 1 | 2 | 91. Talks about killing self | |
| 0 | 1 | 2 | 65. Refuses to talk | 0 | 1 | 2 | 92. Underachieving, not working up to potential | |
| 0 | 1 | 2 | 66. Repeats certain acts over and over; compulsions (describe): _____ | 0 | 1 | 2 | 93. Talks too much | |
| 0 | 1 | 2 | 67. Disrupts class discipline | 0 | 1 | 2 | 94. Teases a lot | |
| 0 | 1 | 2 | 68. Screams a lot | 0 | 1 | 2 | 95. Temper tantrums or hot temper | |
| 0 | 1 | 2 | 69. Secretive, keeps things to self | 0 | 1 | 2 | 96. Seems preoccupied with sex | |
| 0 | 1 | 2 | 70. Sees things that aren't there (describe): _____ | 0 | 1 | 2 | 97. Threatens people | |
| 0 | 1 | 2 | 71. Self-conscious or easily embarrassed | 0 | 1 | 2 | 98. Tardy to school or class | |
| 0 | 1 | 2 | 72. Messy work | 0 | 1 | 2 | 99. Too concerned with neatness or cleanliness | |
| 0 | 1 | 2 | 73. Behaves irresponsibly (describe): _____ | 0 | 1 | 2 | 100. Fails to carry out assigned tasks | |
| 0 | 1 | 2 | 74. Showing off or clowning | 0 | 1 | 2 | 101. Truancy or unexplained absence | |
| 0 | 1 | 2 | 75. Shy or timid | 0 | 1 | 2 | 102. Underactive, slow moving, or lacks energy | |
| 0 | 1 | 2 | 76. Explosive and unpredictable behavior | 0 | 1 | 2 | 103. Unhappy, sad, or depressed | |
| 0 | 1 | 2 | 77. Demands must be met immediately, easily frustrated | 0 | 1 | 2 | 104. Unusually loud | |
| 0 | 1 | 2 | 78. Inattentive, easily distracted | 0 | 1 | 2 | 105. Uses alcohol or drugs (describe): _____ | |
| 0 | 1 | 2 | 79. Speech problem (describe): _____ | 0 | 1 | 2 | 106. Overly anxious to please | |
| 0 | 1 | 2 | 80. Stares blankly | 0 | 1 | 2 | 107. Dislikes school | |
| 0 | 1 | 2 | 81. Feels hurt when criticized | 0 | 1 | 2 | 108. Is afraid of making mistakes | |
| 0 | 1 | 2 | 82. Steals | 0 | 1 | 2 | 109. Whining | |
| 0 | 1 | 2 | 83. Stores up things he/she doesn't need (describe): _____ | 0 | 1 | 2 | 110. Unclean personal appearance | |
| | | | | 0 | 1 | 2 | 111. Withdrawn, doesn't get involved with others | |
| | | | | 0 | 1 | 2 | 112. Worrying | |
| | | | | | | | 113. Please write in any problems the pupil has that were not listed above: | |
| | | | | 0 | 1 | 2 | _____ | |
| | | | | 0 | 1 | 2 | _____ | |
| | | | | 0 | 1 | 2 | _____ | |

APPENDIX B

Cumulative frequency distribution of total problem scores in the normal sample for each gender and age groups 4 - 5 and 6 - 11.

| Total Problem Score | 4-5-year-olds | | 6-11-year-olds | |
|---------------------|-----------------|------------------|-----------------|------------------|
| | Boys (N=114) | Girls (N=130) | Boys (N=351) | Girls (N=410) |
| 15 | 57.0 | 66.9 | 51.0 | 68.8 |
| 16 | 58.8 | 70.8 | 53.8 | 69.3 |
| 17 | 61.4 | 73.8 | 56.1 | 72.7 |
| 18 | 63.2 | 75.4 | 59.0 | 72.9 |
| 19 | 64.0 | 77.7 | 59.5 | 74.4 |
| 20 | 67.5 | 80.0 | 61.3 | 76.6 |
| 21 | 69.3 | 82.3 | 63.8 | 78.5 |
| 22 | 71.9 | 82.3 | 65.5 | 79.5 |
| 23 | 73.7 | 83.1 | 67.0 | 80.2 |
| 24 | 76.3 | 83.8 | 68.4 | 81.2 |
| 25 | 78.1 | 84.6 | 69.8 | 81.5 |
| 26 | 78.1 | 86.9 | 71.5 | 82.2 |
| 27 | 78.9 | 86.9 | 73.5 | 84.1 |
| 28 | 78.9 | 87.7 | 74.9 | 84.4 |
| 29 | 81.6 | 87.7 | 75.8 | 85.1 |
| 30 | 82.5 | 88.5 | 76.9 | 86.1 |
| 31 | 84.2 | 90.0 | 77.2 | 86.6 |
| 32 | 84.2 | 90.0 | 78.3 | 87.3 |
| 33 | 85.1 | 90.0 | 80.1 | 87.8 |
| 34 | 86.0 | 90.8 | 81.5 | 87.8 |
| 35 | 86.8 | 90.8 | 82.3 | 88.8 |
| 36 | 87.7 | 91.5 | 82.6 | 89.8 |
| 37 | 87.7 | 91.5 | 83.2 | 90.5 |
| 38 | 87.7 | 91.5 | 84.9 | 91.5 |
| 39 | 87.7 | 91.5 | 85.2 | 92.2 |
| 40 | 88.6 | 91.5 | 85.8 | 92.4 |
| 41 | 88.6 | 91.5 | 86.6 | 92.4 |
| 42 | 90.4 | 91.5 | 87.2 | 92.7 |
| 43 | 91.2 | 93.1 | 88.0 | 92.7 |
| 44 | 92.1 | 93.1 | 88.3 | 93.2 |
| 45 | 93.9 | 93.8 | 88.9 | 93.2 |
| 46 | 93.9 | 93.8 | 89.2 | 93.7 |
| 47 | 93.9 | 93.8 | 89.7 | 93.9 |
| 48 | 94.7 | 93.8 | 89.7 | 94.1 |
| 49 | 94.7 | 94.6 | 90.3 | 95.4 |
| 50 | 95.6 | 94.6 | 90.9 | 95.6 |

APPENDIX C

Percentage of variance accounted for by significant ($p < .01$) effects of gender and age in ANCOVAs of behavior problems.

| Item | Gender | Age | SES |
|-------------------------------------|------------------|------------------|-----------------|
| 1. Acts too young | 1 ^M | | |
| 2. Hums or makes other odd noises | 3 ^M | | |
| 3. Argues a lot | 1 ^M | | |
| 4. Fails to finish things | 1 ^M | | |
| 5. Behaves like opposite sex | | | |
| 6. Defiant, talks back to staff | <1 ^M | | |
| 7. Bragging | 12 ^M | | |
| 8. Can't concentrate | 3 ^M | | |
| 9. Obsessions | | | |
| 10. Hyperactive | 3 ^M | | |
| 11. Too dependent | | | <1 |
| 12. Lonely | | | |
| 13. Confused | | 2 ⁰ | <1 [*] |
| 14. Cries a lot | | | |
| 15. Fidgets | 2 ^M | | |
| 16. Cruelty | <1 ^{M*} | | |
| 17. Day-dreams | | 1 ^{NL} | |
| 18. Harms self | | | |
| 19. Demands attention | | | |
| 20. Destroys own things | 1 ^M | | |
| 21. Destroys others' things | | | |
| 22. Difficulty following directions | <1 ^M | | |
| 23. Disobedient at school | 1 ^M | | <1 [*] |
| 24. Disturbs other pupils | 3 ^M | 1 ⁰ | <1 [*] |
| 25. Poor peer relations | | 1 ^{NL*} | |
| 26. Lacks guilt | 2 ^M | | |
| 27. Easily jealous | 1 ^{F*} | | |
| 28. Eats non-food | | | |
| 29. Fears | | | |
| 30. Fears school | | | |
| 31. Fears impulses | 1 ^F | | |
| 32. Needs to be perfect | | | 2 ^U |

| Item | Gender | Age | SES | Item | Gender | Age | SES |
|---|-------------------|-------------------|-------------------|--|-----------------|------------------|-----------------|
| 33. Feels unloved | | 2 ⁰ | | 58. Picking | | <1 ^H | <1 [*] |
| 34. Feels persecuted | | 2 ⁰ | | 59. Sleeps in class | | 1 ^{NL} | |
| 35. Feels worthless | | 2 ⁰ | | 60. Apathetic or unmotivated | | 3 ^M | |
| 36. Accident prone | 1 ^M | | 1 | 61. Poor schoolwork | 1 ^M | 2 ⁰ | 2 |
| 37. Fighting | 6 ^M | <1 ^Y * | | 62. Clumsy | 3 ^M | | <1 |
| 38. Is teased | | 1 ⁰ | <1 [*] | 63. Prefers older children | | 1 ^Y * | |
| 39. Hangs around with children who get in trouble | | | | 64. Prefers younger children | | | |
| 40. Hears things that aren't there | | | | 65. Refuses to talk | | | |
| 41. Impulsive | 4 ^M | | 1 | 66. Compulsions | | | |
| 42. Likes to be alone | | | | 67. Disrupts class discipline | 1 ^M | | |
| 43. Lying or cheating | 1 ^M | | | 68. Screams a lot | 1 ^M | | |
| 44. Bites fingernails | | | | 69. Secretive | | 1 ⁰ | |
| 45. Nervous | <1 ^M | | 1 ^{NL} * | 70. Sees things that aren't there | | | |
| 46. Nervous movements | <1 ^M * | | | 71. Selfconscious | | | |
| 47. Overconforms to rules | | | | 72. Messy work | 6 ^M | 2 ⁰ | |
| 48. Not liked | | 2 ⁰ | | 73. Behaves irresponsibly | 1 ^M | | |
| 49. Difficulty learning | <1 ^M | 4 ⁰ | 1 | 74. Showing off | 7 ^M | | |
| 50. Too fearful or anxious | | | | 75. Shy or timid | | | |
| 51. Feels dizzy | | | | 76. Explosive and unpredictable behavior | 1 ^M | | |
| 52. Feels too guilty | | | | 77. Easily frustrated | | | |
| 53. Talks out of turn | 2 ^M | | | 78. Inattentive | 3 ^M | | <1 |
| 54. Overtired | | 2 ⁰ | | 79. Speech problem | 1 ^M | 1 ^{NL} | |
| 55. Overweight | | | | 80. Stares blankly | | | |
| 56a Aches or pains | | 2 ⁰ | | 81. Feels hurt when criticized | <1 ^M | | |
| 56b Headaches | | | | 82. Steals | | | |
| 56c Nausea, feels sick | | | | 83. Stores up unneeded things | <1 ^M | | |
| 56d Eye problems | | | | 84. Strange behavior | | | |
| 56e Skin problems | | 1 ⁰ * | | 85. Strange ideas | | | |
| 56f Stomachaches or cramps | | | | 86. Stubborn | | | |
| 56g Vomiting | | | | 87. Moody | | | |
| 56h Other | | | | 88. Sulks a lot | | | |
| 57. Physically attacks people | 2 ^M | | | 89. Suspicious | | | |
| | | | | 90. Swearing | 1 ^M | | 1 |

| Item | Gender | Age | SES |
|--|------------------|----------------|-----|
| 91. Suicidal talk | | | |
| 92. Underachieving | 2 ^M | | |
| 93. Talks too much | | | |
| 94. Teases a lot | 2 ^M | | <1 |
| 95. Temper tantrums | 2 ^M | | |
| 96. Sexual preoccupation | | 1 ⁰ | <1 |
| 97. Threatens people | <1 ^M | | |
| 98. Tardy to school or class | | | |
| 99. Too concerned with neatness | <1 ^{F*} | | |
| 100. Fails to carry out assigned tasks | <1 ^M | | |
| 101. Truancy | | | |
| 102. Underactive | | | |
| 103. Unhappy, sad or depressed | | | |
| 104. Unusually loud | 3 ^M | | |
| 105. Alcohol or drugs | | | |
| 106. Overly anxious to please | | | |
| 107. Dislikes school | | | |
| 108. Is afraid of making mistakes | | | |
| 109. Whining | | | <1 |
| 110. Unclean personal appearance | | | <1 |
| 111. Withdrawn | | | |
| 112. Worrying | | | |
| Total behavior problem score | 3 ^M | | 1 |

M = higher scores for males

F = higher scores for females

0 = higher scores for older children

Y = higher scores for younger children

NL = nonlinear effect of age

U = higher score for upper SES; all other effects show higher scores for lower SES

* Indicates for each of the significant main effects those 5 effects having the smallest F values (see text).

REFERENCES

1. Lapouse R, Monk MA. An epidemiological study of behavior characteristics in children. *Am J Public Health* 1958, 48, 1134-1144.
2. Verhulst FC, Akkerhuis GW, Althaus M. Mental health in Dutch children: (I) a cross-cultural comparison. *Acta Psychiatrica Scandinavica*, 1985, 72, suppl. no 323.
3. Achenbach TM, Edelbrock CS. Behavioral problems and competencies reported by parents of normal and disturbed children aged four through sixteen. *Monogr Soc Res Child Dev* 1981, 188.
4. Achenbach TM, Edelbrock CS. Manual for the child behavior checklist and revised child behavior profiles. Burlington: University of Vermont, Department of Psychiatry, 1983.
5. Achenbach TM, Verhulst FC, Baron D, Akkerhuis GW. Epidemiologic comparisons of American and Dutch children: I Behavioral-emotional problems reported by parents for ages 4-16, 1986 (submitted).
6. Rutter M, Graham P. The reliability and validity of the psychiatric assessment of the child: I. Interview with the child. *Br J Psychiatry* 1968, 114, 563-579.
7. Mitchell S, Sheperd M. A comparative study of children's behaviour at home and at school. *Br J Educ Psychol* 1966, 36, 248-254.
8. Gould Schwartz M, Wunsch-Hitzig R, Dohrenwend B. Estimating the prevalence of childhood psychopathology: a critical review. *J Am Acad Child Psychiatry* 1981, 20, 462-476.
9. Earls F. The prevalence of behavior problems in 3-year-old-children: comparison of the reports of mothers and fathers. *J Am Acad Child Psychiatry* 1980, 19, 439-452.
10. Achenbach TM, McConaughy SH, Howell CT. Child/Adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity, 1986, *J Cons Clin Psychol* (accepted).
11. McGee R, Silva PA, Williams S. Behaviour problems in a population of seven-year-old children; prevalence, stability and types of disorder: a research report. *J Child Psychol Psychiatry* 1984, 25, 251-259.
12. Sprague RL, Cohen MN, Eichlseder W. Are there hyperactive children in Europe and the South Pacific? In: Halliday R. *The hyperactive child: fact, fiction, and fantasy*. Symposium presented at Am Psychol Assoc. San Francisco, 1977.
13. Achenbach TM, Edelbrock CS. Manual for the Teacher's Report Form and teacher version of the Child Behavior Profile, Burlington: University of Vermont, Department of Psychiatry, 1986.
14. Conners CK. A teacher rating scale for use in drug studies with children. *Am J Psychiat*, 1969, 126, 152-156.
15. Quay HC, Peterson DR. Interim manual for the revised behavior problem checklist. Applied social sciences. Coral Gables: University of Miami, 1983.

ACKNOWLEDGEMENTS

This project would not have been possible without the continuous support from Prof. Thomas M. Achenbach, who advised us from the very beginning of our study until the last revision of this report. We are also grateful to the head of our department Prof. J.A.R. Sanders-Woudstra, who from the beginning stimulated the idea that child psychiatric epidemiological research needs to be carried out.

Of the many people who worked on this project, we are especially indebted to Annegret de Bel and Jaantje Prince for their skilful administrative work, to Astrid Melker who typed the manuscript and to John de Kuyper for his work on the figures. We wish to thank the many teachers who gave their time to fill in the checklists.