Problem behavior in Dutch preschoolers

Probleemgedrag bij Nederlandse peuters en kleuters

Proefschrift

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de Rector Magnificus
Prof. Dr. P.W.C. Akkermans M.Lit.
en volgens besluit van het College van Dekanen.

De openbare verdediging zal plaatsvinden op
woensdag 6 oktober 1993 om 13.45 uur

door

Johannes Maria Koot

geboren te Valkenswaard.

Promotiecommissie

Promotor:	Prof. Dr. F.C. Verhulst
Overige leden:	Prof. Dr. H. van Engeland
	Prof. Dr. R. Giel
	Prof. Dr. J. Passchier

De zee

JUDITH HERZBERG

De zee kun je horen
met je handen voor je oren,
in een kokkel,
in een mosterdpotje,
of aan zee.

The studies reported in this thesis were performed at the department of Child and Adolescent Psychiatry, Sophia Children's Hospital / Erasmus University Rotterdam, The Netherlands.

This research was supported by the Dutch National Programme for Stimulation of Health Research (SGO), the Dutch Ministry of Well-being, Health, and Cultural Affairs (WVC), and the Rotterdam Foundation for Autism and Congenital Contact Disorders (SAR).

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Koot, Johannes Maria

Problem behavior in Dutch preschoolers / Johannes Maria Koot. - [S.l. : s.n]
Thesis Rotterdam. - With ref. - With summary in Dutch.
ISBN 90-9006482-6
Subject headings: problem behavior; preschoolers.

Cover: Gerard Koot

Contents

Abbreviations

AGFI	Adjusted Goodness of Fit Index
BCL	Behaviour CheckList
BSQ	Behaviour Screening Questionnaire
CBCL/2-3	Child Behavior Checklist for Ages 2-3
CBCL/4-18	Child Behavior Checklist for Ages 4-18
CFA	Confirmatory Factor Analysis
DOTS-R	Dimensions of Temperament Survey-Revised
DSM-III-R	Diagnostic and Statistical Manual, third revised edition
EFA	Exploratory Factor Analysis
EPC	Expected Parameter Change
FIF	File Information Form
GFI	Goodness of Fit Index
ICD-10	International Classification of Diseases, tenth revision
LEQ	Life Events Questionnaire
LSI	Language Screening Instrument
MCDI	Minnesota Child Development Inventory
MI	Modification Index
MSCA	McCarthy Scales of Children's Abilities
NOSP	Nijmegen Observation Scales for Preschoolers
NPSI	Nijmegen Parenting Stress Index
OR	Odds Ratio
PBQ	Preschool Behavior Questionnaire
PBCL	Preschool Behaviour Checklist
PCA	Principal Components Analysis
PFA	Principal Factor Analysis
RMR	Root of Mean squared Residuals
SES	SocioEconomic Status
TRF	Teacher's Report Form
YSR	Youth Self Report

CHAPTER 1

Introduction

1.1 Background

In the last two decades the interest in the psychosocial development of young children has increased considerably. Not only have new theories and fields of research on the social and emotional development emerged (e.g., attachment theory and research; theories of self-regulation), also behavioral and emotional problems specific to this age range have received more attention than before (e.g., Campbell, 1990; Richman & Lansdown, 1988; Trad, 1988). However, basic issues in the study of developmental psychopathology of children younger than 4-5 years remain relatively unaddressed. Developmental psychopathology studies developmentally and normatively deviant behaviors and may include research areas such as assessment, taxonomy, epidemiology, developmental course, prognosis, etiology, and treatment, which are both hierarchical and recursively related to each other (Achenbach, 1990). In the study of psychopathology in very young children, even the basic issues of assessment, taxonomy, and epidemiology of deviant behavior and development have only been scantily addressed. While for older age groups numerous efforts have been undertaken to develop standardized assessment procedures and taxonomic systems of psychopathology (e.g., Achenbach, 1993; Rutter, Tuma, & Lann, 1988), and a considerable number of prevalence studies have been performed involving over 123,000 children (see Verhulst & Koot, 1992, for an overview), firm data on the prevalence and correlates of symptoms, and on the validity of differentiated syndromes of psychopathology in children below age 4 are still lacking.

One reason for the lack of information on psychopathology in young preschool children may be that it is difficult to differentiate between behaviors that can be regarded normal for this age and behaviors that can be considered deviant or pathological. Generally, a consistent picture of behavior in the young preschool child (2-3 years) is difficult to obtain from either parents, who at this age are the primary source of information, or other informants such as clinicians, because in young children normal functioning may be characterized by fairly extensive fluctuations in behavior from day to day. Further, many behaviors considered problematic in older children, like defiance, anxiety in new situations, or lack of emotional and behavioral control, may be part of normal development in younger children. Finally, different informants seeing the preschool child under different conditions and having different relations to the child often disagree on the presence and severity of the child's problem

behaviors. Each informant may provide valid, though different data. Not surprisingly, no coherent taxonomy exists that might be used to describe behavioral and emotional problems in preschool children. For example, DSM-III-R (APA, 1987) diagnostic categories, such as oppositional disorder or separation anxiety disorder, may have little validity in preschool children, because many otherwise normally functioning children will qualify for these diagnoses. Children's problem behaviors generally involve quantitative variations on behavior that most preschool children display to some degree during their development.

Given the problems in defining psychopathology in young children, it may not be surprising that studies on the occurrence of psychopathology in preschool children have yielded widely varying results. The reported prevalence based on parental reports ranged from 6% to 50%. This is so because prevalence rates were derived from diverse and often unstandardized assessment procedures. A further limitation of epidemiological studies of problem behavior in young preschoolers is that most of these used unrepresentative samples, and many studies differed in their criteria for deviance, i.e. simple percentages of reported behavior, and cutoff points based on either percentile scores or on ratings made by psychiatrists. Of the epidemiological studies of problem behavior in preschoolers, the study by Richman, Stevenson, and Graham (1975, 1982) was the most comprehensive, involving a sample of 705 3-year-olds in an outer London borough, of whom about 200 were followed-up until age 8 years, and including the study of many important correlates of young children's behavior. However, this study used a 12-item interview-based questionnaire (the Behaviour Screening Questionnaire; BSQ) to obtain parents' reports of problem behavior. And although the BSQ is reported to be a reliable and valid screening instrument to detect preschool children with behavioral problems, it does not give detailed reports on children's problem behavior. The main distinction, obtained by factor-analysis, was between two broad-band dimensions concerning "conduct-restless" versus "emotional-miserable" syndromes. Furthermore, the cutoff score on the BSQ that creates normal and abnormal ranges has been based on psychiatrists' ratings of the severity of behavioral disorder from the same institute, which may only partially converge with ratings obtained from other observers.

In the Netherlands, information on the prevalence of problem behavior in young preschoolers was completely absent, and a validated instrument to provide such information was not available, although a translation of the BSQ had been used in at least two studies (Swets-Gronert, 1986; Verhulst & Akkerhuis, 1983). In 1988, the Dutch National Committee for the Early Detection of Developmental Disturbance (Landelijke Commissie VTO, 1988, pp.76-80) still had to point at the existing need for behavior checklists and behavioral observation ratings to detect signs of general and specific developmental delays and disturbances in children aged 0-6 years.

Since 1983 the research programme of the Department of Child and Adolescent Psychiatry of the Sophia Children's Hospital and Erasmus University Rotterdam has been aimed at the development and standardization of checklists to collect data on a broad range of behavioral and emotional problems in children and adolescents. This was performed in close cooperation with T.M. Achenbach and coworkers (University of Vermont), who

developed checklists for the description of problems that are of clinical relevance for youth in the age range of 4-18 years to be filled out by parents, teachers, and youths aged 11 years and up (Achenbach, 1991a, b, c, d). These instruments have consistently yielded eight syndromes of problem behavior that appear to have considerable discriminative and predictive validity. For each of these instruments Dutch normative data were collected and their discriminative and cross-cultural validity was established.

Since 1986 a downward extension of the checklist became available to be completed by parents and caregivers of children ages 2-3 years, the Child Behavior Checklist for Ages 2-3 (CBCL/2-3), providing information of six syndromes of psychopathology in young children. Research in the United States (Achenbach, Edelbrock, & Howell, 1987) indicated good reliability and discriminative validity of the instrument. The instrument was soon translated into Dutch (by F.C. Verhulst) and used in clinical settings. However, normative and psychometric data for the instrument were still absent. The present study was designed to provide information on the usefulness of the instrument in the Dutch situation. The main aims were to assess the psychometric qualities of the instrument (internal structure, reliability, concurrent, criterion-related, and predictive validity), and to provide data on the prevalence and correlates of problem behavior in Dutch referred and nonreferred preschoolers ages 2-3 using the instrument.

1.2 Aims of the study

The present study was an attempt to contribute to the standardized assessment and the study of prevalence of problem behavior in children 2- to 3-years old by means of empirically derived rating scales. The basic questions were:

1. What are the psychometric characteristics of the Dutch version of the Child Behavior Checklist for Ages 2-3 (CBCL/2-3)?
2. What is the prevalence of problem behavior in Dutch toddlers aged 2-3 years in the community, and in a broadly defined clinical sample as measured by the CBCL/2-3?
3. How are CBCL/2-3 scores in both samples related to demographic, family, and child characteristics, and in the clinical sample also to clinically relevant variables, such as referral complaints and psychiatric diagnosis?
4. How predictive are CBCL/2-3 scores of behavioral/emotional problems at home and at school over a 2-year interval, as measured by parents' and teachers' standardized ratings of problem behavior, and reports of significant problems at home and school?

1.3 Structure of the report

Chapter 2 gives an overview of efforts to standardize the assessment of problem behavior in preschoolers, and the results of relevant studies on its prevalence and longitudinal course are discussed. The samples and methods of data collection used in this study are described in Chapter 3.

The first aim of the study was to obtain psychometric data on the Dutch version of a standardized checklist to obtain parent ratings of toddlers' problem behavior, i.e. the Child Behavior Checklist for Ages 2-3 (CBCL/2-3) originally developed by Achenbach, Edelbrock and Howell (1987). This was done by studying the factor structure in three Dutch samples of 2-3-year-old children: a sample of children from the general population, a sample of children referred to mental health settings, and a large sample of twin pairs; by studying the internal consistency and test-retest reliability of the checklist; by studying its relation to other instruments designed to measure similar and dissimilar phenomena; and by studying its power to differentiate between normal children and children with acknowledged behavioral/emotional problems. The results of this part of the study will be presented in Chapters 4 and 5.

The second and third aim of the study were to provide prevalence rates, and to study correlates of problem behavior in toddlers in the general Dutch population and in toddlers referred to mental health settings using the CBCL/2-3. This was done by obtaining parent ratings of problem behavior in a sample of 420 2-3-year-olds drawn randomly from the general population of the Dutch province of Zuid-Holland, and in a sample of 458 children of the same age referred to one child guidance service, four community mental health centers, four child psychiatric outpatient clinics, and three medical daycare centers. Correlates included in both samples: the composition and social position of the child's family of origin, child health, parental psychopathology, and parenting characteristics. In the clinical sample, clinical correlates of parents' problem ratings such as referral complaints and diagnosis were studied in addition. Results from these two prevalence studies will be described in Chapters 6 and 7.

Chapter 8 regards the fourth aim of the study, which was to assess the stability and predictive validity of toddlers' problem behavior as measured by the CBCL/2-3. This was done in a prospective follow-up study of the community sample two years after initial assessment. The stability and change of parents' reports of children's problem behavior was assessed over the two-year period. Further, it was determined how well parents' initial reports of problem behavior predicted parents' and teachers' reports of problem behavior and signs of significant problems, over and above family and child characteristics and intervening stressful experiences.

In Chapter 9 conclusions and implications of the study's results will be discussed.

CHAPTER 2

Problem behavior in young preschoolers

2.1 Introduction

Most clinicians will agree that a behavioral disorder in young children must include a pattern of symptoms, that has been troublesome for some time to the child and/or others, occurs in more than one situation, is relatively severe, and is likely to impede the child's performance in age-appropriate developmental tasks (cf. Campbell, 1990). Not the sheer presence of problem behaviors differentiates normal from deviant, but the frequency, intensity, chronicity, constellation, and (social) context of the behavior (cf. Rutter, Tizard, & Whitmore, 1970). However, for young preschoolers these guidelines have nor led to the development of a generally accepted validated system for the description and classification of problem behaviors. Such a taxonomic system is important in clinical practice to assign a proper diagnosis to a case, to be able to communicate about a child's behavior, and to assign cases to treatment conditions. A taxonomy is equally important in research on the origin, correlates, course, and intervention in children's problem behavior.

In young preschoolers it is difficult to differentiate between behaviors that can be regarded normal for a certain age and behaviors that may be indicative of psychopathology. This has several reasons.

First, a consistent picture of behavior in the young preschooler is difficult to obtain from either parents or other informants such as clinicians, because in young children normal functioning may be characterized by fairly extensive fluctuations in behavior from day to day.

Second, many isolated behaviors considered problematic in older children are very common in nonreferred children of young preschool age (e.g., Achenbach, 1992; Macfarlane, Allen, & Honzik, 1962; Richman, Stevenson, & Graham, 1982), and may be part of normal development in young preschoolers. Thus, bedwetting, stranger anxiety, or separation distress will not be considered problematic among most 2-year-olds. Defiant or negativistic behavior, being of concern among school-aged children, may be regarded quite common among toddlers seeking to establish autonomy during the separation-individuation process (Mahler, Pine, & Bergman, 1975; Wenar, 1982). Likewise, given the level of cognitive development of most toddlers, aggressive encounters between 2- and 3-year-olds may more often be interpreted as rough instead of aggressive. The toddler is not yet able to

take the viewpoint of others into account, and thus not connect certain behaviors (e.g., hitting) with their effect on another person (i.e., inflicting pain).

Third, different informants seeing the preschooler under different conditions and having different relations to the child often disagree on the presence and severity of the child's problem behaviors. Each informant may provide valid, though different data. This phenomenon has been observed for both older and younger children using reports of behavioral / emotional problems (Achenbach, McConaughy, & Howell, 1987), as well as temperament ratings (e.g., Rothbart & Goldsmith, 1985).

Fourth, no coherent taxonomy exists that might be used to describe behavioral / emotional problems in young preschoolers. For example, DSM-III-R (APA, 1987) and ICD-10 (WHO, 1989) diagnostic categories, such as oppositional defiant disorder or separation anxiety disorder, may have little validity in preschool children, because many otherwise normally functioning children will qualify for these diagnoses. With the possible exception of infantile autism, most problem behaviors in young preschoolers do not form clearcut diagnostic categories, like those included in these classification systems. Instead, these problem behaviors generally involve quantitative variations on behavior that most young children display to some degree during their development.

In an attempt to provide conceptual clarity in the description of young children's potentially problematic behaviors, Campbell (1990) suggested the terms (1) annoying behavior, (2) problem behavior, and (3) symptomatic behavior. Annoying or worrisome behavior refers to typical or age-appropriate behavior that may concern some parents. Problem behavior may be used to indicate an exaggeration in the frequency and/or intensity of typical behavior to an upsetting degree. Symptomatic behavior or symptoms may be utilized to designate a problem of probable clinical significance. Although potentially useful, as Campbell noted, it is difficult, if not impossible, to clearly differentiate between these degrees of troublesome behavior. In this dissertation we prefer the term problem behavior to indicate those behaviors, that may or may not be clinically significant, without implying the distinction proposed by Campbell. One of the aims of the present research was to provide information that enables valid distinctions between different degrees of troublesome behavior.

2.2 Approaches to the assessment of psychopathology in children

In searching for methods to define and assess behavioral and emotional disorders in children it should be acknowledged, that several approaches to the assessment and classification of problem behavior exist that emphasize different criteria of psychopathology. Two dominant approaches have been the clinical consensus approach and the psychometric approach. More recently, the developmental deviance approach has been added. Each of these approaches will be discussed briefly.

Clinical consensus approach

This approach has originated from the medical assessment tradition, that is characterized by the use of clinical interview techniques that rely on the experience and skill of the clinician

in arriving at an accurate diagnosis. Clinical consensus is used to describe syndromes and their associated clusters of symptoms. Clinical classification systems that have been shaped by this approach, and that cover the majority of child psychiatric conditions are the Diagnostic and Statistical Manuals of the American Psychiatric Association (e.g., DSM-III-R; APA, 1987) and the International Classification of Diseases of the World Health Organization (e.g., ICD-10; WHO, 1989). Although they have been welcomed because they enhanced the standardization of diagnostic formulations, these systems have some drawbacks that make them less useful for the study of problem behavior in young preschoolers.

One basic problem is that the criteria of the categories are rather loosely defined, and that standardized assessment procedures for identifying criteria are lacking. For example, although for several categories reference is made to appropriateness of behaviors for a certain age, it is left to the judgment of the clinician whether specific behaviors or frequency and intensity of behaviors should be considered deviant or not. A second problem is that the diagnostic categories of these systems are based on the presence of a fixed number of symptoms, of which it has not empirically been established that they form a pattern. The descriptions of these categories have been formulated by a process of negotiation rather than empirical evidence. A third problem may be that the categorical definitions are relatively insensitive to developmental changes in occurence and covariation of symptoms. This makes the systems less useful for longitudinal studies.

Psychometric approach

The psychometric approach to definition of child psychopathology originated from the experience in psychology with the assessment of psychological traits, such as intelligence or personality, via standardized tests and scales.

This approach is characterized, first, by the use of standardized assessment procedures, usually in the form of rating scales. Several rating scales have been developed comprising a broad range of problem behaviors, to be scored by parents, teachers, or children themselves. Parents are usually the most important sources of information on young children's behavior, because they are familiar with their child's behavior across time and across many situations. If the child visits a daycare setting or a nursery school, daycare personnel or nursery school teachers also may provide useful information.

Second, quantitative scores for syndrome scales are used. Multivariate statistical techniques have been employed to construct scales with statistically related items that are the empirical counterparts of clinical syndromes (see Achenbach, 1985; Quay, 1986; Barkley, 1988, for an overview). These scales, when used as continuous measures, retain more statistical information than present versus absent categories, and are sensitive to changes over time (e.g., Verhulst, Koot, & Berden, 1990). Scores are used to indicate how an individual child compares with reference groups. Quantitative scores derived from standardized assessment procedures are compared with the scores for normative samples of peers of the same sex and age. Baseline data, obtained for normative and clinical samples using rating scales or questionnaires, make it possible to compare an individual child's score with scores reported for clinical and non-clinical agemates. This may aid the clinician in forming a

comprehensive picture of the child's problem behavior. Researchers, too, may benefit from this approach, making comparisons within and between groups possible. Especially large-scale epidemiological research is practically impossible without the use of rating scales for assessing problem behavior, provided these procedures are reliable and valid.

Finally, empirical syndromes can be used to arrive at present versus absent categories of psychopathology by imposing cutoff scores. In clinical practice the categorical approach is often preferred, because this may facilitate decisions that need to be made regarding management or intervention. Much information would be lost if cutoff points were imposed arbitrarily to arrive at present versus absent categories. However, the usefulness of particular cutoff points can be tested empirically, for example by computing Receiver Operating Characteristics (e.g., see Achenbach, 1992; Verhulst & Koot, 1992).

Developmental deviation approach

The developmental deviation approach was formulated to define child psychopathology more developmentally. As Sroufe (1989) remarks, the clinical consensus approach, as represented by DSM has contributed to the developmental aspects of psychopathology little more than to mention age of onset of disorders. In his view, the effort of the psychometric approach has been to select a series of problem behaviors and symptoms, which might apply across much of childhood rather than keying to age-appropriate issues.

Using Anna Freud's (1965) work as a starting point, the developmental deviation approach underscores (1) a focus on age appropriate developmental issues, such as the establishment of an effective attachment relationship (from 6 to 12 months of age) or autonomy (18-30 months) (see Sroufe, 1979, 1989; Sroufe & Rutter, 1984); (2) assessment of current behavior in terms of those issues; and (3) analysis of the profile of adaptation of a child cutting across aspects of development and defined in qualitative rather than quantitative terms. Individuals manifesting developmental deviation are in the words of Sroufe (1989) "... those whose manner of dealing with the set of issues for a given period of development puts them at a disadvantage for dealing with the issues of a subsequent period. (...) The judgment of whether they are now or will likely later be disordered depends on the particular nature of the deviation, as does the particular set of disorders likely to be manifest" (p. 19). Some authors define potential relationships between specific psychopathological disorders and specific developmental tasks (e.g., see Garber, 1984).

Although this approach seems promising as a way to assess the meaning of problem behaviors from a developmental perspective, the description of developmental issues and patterns of adaptation across phases and aspects of development has only started recently. Examples of developed research on developmental tasks and patterns of adaptation and deviance regard the early caregiver-child attachment and peer relationships at school age. Issues which may be salient for 2-3 years olds, such as the establishment of individuation or autonomy and flexible self-control have not yet been studied from this perspective. Furthermore, it seems difficult to disentangle behaviors indicating that the child is dealing with a developmental task from those indicating the presence of developmental deviance. Finally, although the approach seems conceptually attractive it relieves the researcher not

from the responsibility to develop measures of deviance which are both psychometrically sound and may be useful for both clinical and research purposes.

We may conclude that at present neither the clinical concensus approach nor the developmental deviance approach provide a proper basis for the assessment of problem behavior in young preschoolers. By contrast, the psychometric apporach offers a rich tradition of assessment of problem behavior at different ages. Efforts to develop standardized procedures for the assessment of problem behaviors in preschoolers will be discussed in the following sections.

2.3 Standardized assessment of problem behavior in preschoolers

Numerous parent, teacher, as well as (for adolescents) self-report rating scales have been developed to obtain behavior ratings for children aged 4-18 (e.g., Achenbach & Edelbrock, 1983, 1986, 1987; Goyette, Conners, & Ulrich, 1978; Quay & Peterson, 1983; see also Barkley, 1988). For these age groups multivariate analyses of behavioral and emotional problems yielded multiple syndromes, that have been replicated for various sex and age groups across samples and informants (e.g., Achenbach 1991a; Achenbach, Conners, Quay, Verhulst, & Howell, 1989). Further, higher order groupings have been identified across studies and instruments for which Internalizing and Externalizing (Achenbach, 1966) have now become generally accepted labels (Cicchetti & Toth, 1991). However, only few efforts have been devoted to the standardized assessment of problem behaviors in 2- and 3-year olds. This reflects (1) a general paucity of research on psychopathology in the 2- to 3-year age period; and (2) difficulties in defining criteria for deviance and a lack of differentiated concepts of psychopathology for these ages (cf. Achenbach, Edelbrock, & Howell, 1987).

The next sections give a description of instruments with known psychometric properties for obtaining parent and teacher ratings of a broad range of problem behavior developed specifically for an age range including ages 2 and 3 years. An overview of the instruments is given in Table 2.1. Rating scales which are of limited value, due to a lack of psychometric and normative data (e.g., Bell, Waldrop, & Weller, 1972; Eisenberg, Landowne, Wilmer, & Iamber, 1962), developed only for older preschoolers (ages 5 and 6; e.g., Becker, 1960; Caldwell & Pianta, 1990; Pianta, 1987; Reinherz & Gracey, 1983), or for a very limited population (e.g., Brown, 1960) have been omitted.

2.3.1 Parent rating scales and questionnaires

Behaviour Screening Questionnaire (BSQ)

The BSQ (Richman & Graham, 1971) was developed to assess problems in 3-year-olds in a standardized way. It consists of questions that are administered by trained interviewers to mothers whose reports about their children are scored on 3- and 4-step scales for the following 12 problem areas: eating, sleeping, soiling, activity, concentration, relationship with sibs and peers, dependency/attention seeking, control, tempers, mood, worries, and

Table 2.1a Broad-band rating scales for assessing problem behavior in preschool children: Descriptive characteristics

Author, Date	Rating Scale	Age range (years)	Informant	Respons Format	Number of Items	Factors/ Scales
Parent Ratings						
Richman & Graham, 1971	Behavioural Screening Questionnaire	3	Parent	3-step	12	Total BSQ
Richman, 1977	Behaviour Checklist	3	Parent	3-step	19	Total BCL
Achenbach, 1992 (First version: Achenbach, Edelbrock, & Howell, 1987)	Child Behavior Checklist for Ages 2-3	2-3	Parent or Parent-surrogate	3-step	99	1. Anxious/depressed 2. Withdrawn 3. Sleep problems 4. Somatic problems 5. Aggressive behavior 6. Destructive behavior 7. Internalizing 8. Externalizing
Teacher Ratings						
Kohn & Rosman, 1972	Symptom Checklist	3-6	Teacher or Teacher-aid	3-step	49	1. Apathy-withdrawal 2. Anger-defiance
Behar & Stringfield, 1974	Preschool Behavior Questionnaire	3-6	Preschool teacher	3-step	30	1. Hostile-aggressive 2. Anxious-fearful 3. Hyperactive-distractible
McGuire & Richman, 1986	Preschool Behaviour Checklist	2-5	Preschool staff	3-step	22	1. Conduct/restless/aggressive 2. Isolated/immature 3. Emotional/miserable

Table 2.1b Broad-band rating scales for assessing problem behavior in preschool children: Reliability and validity

Author, Date	Reliability: Test-retest	Reliability: Internal Consistency	Validity: Criterion Related	Validity: Construct Related	Other characteristics
Parent Ratings					
Richman & Graham, 1971	1 wk: r = .77	----	Discriminates between normal and deviant by clinical judgment.	----	----
Richman, 1977	4-wk: r = .81	----	Discriminates between normal and deviant by clinical judgment.	----	----
Achenbach, 1992	1 wk, 1 yr: mean r = .85, .64	Mean alpha = .81 (scales 1-6); .91 (scales 7-8)	All scales discriminate well between children referred for mental health services and nonreferred children.	Moderate to high correlations with similar scale from the CBCL/4-16 2-7 years later. High correlation between CBCL/2-3 and BCL (Spiker e.a., 1992).	Interparent agreement: mean r = .63 and .60 for ages 2 and 3, respectively
Teacher Ratings					
Kohn & Rosman, 1972	6, 12, 18 mo: Mean r = .67, .49, and .41	----	Differentiated well and moderately well from poorly functioning, remedial, therapy, day treatment and mental hospital groups. Sensitivity healthy vs. disturbed: boys .67, girls .59 (Kohn & Rosman, 1973).	Factors related .16-.48 with corresponding factors on Peterson Problem Checklist over 12-18 mo. interval.	Agreement between teachers: mean r = .69.
Behar & Stringfield, 1974	3-4 mo: mean r = .60-.94	----	Discriminates well between normals and children from preschools involved in early intervention.	A two- instead of a three-component structure ("Aggressive-hyperactive"; "Anxious-withdrawn") has been found consistently for other samples (Fowler & Park, 1979; Hoge et al., 1985; Tremblay et al., 1987).	Interrater agreement: mean r = .77.
McGuire & Richman, 1986	----	Split-half r = .83 Alpha = .83	Differentiates between psychiatric day centre and nursery samples. Relation to observations/interviews: 33% / 30% false positives, 4% / 2% false negatives. Discriminates between psychiatric day centre and nursery sample.	Relation with PBQ: r = .89, cutoff agreement 81%. Chinese factor structure similar to original (Luk et al., 1991).	Interrater score agreement: r = .68; problem/no problem agreement 83%; cut-off agreement 38%.

Note: Findings reported were obtained from the first publication on the instrument unless stated otherwise.

fears. The scores (0-3) on each area are summed to a total score. Richman and Graham (1971) reported a test-retest correlation of .77 between two interviewers repeating the BSQ over a 1-year interval. A score of 10 or higher on the BSQ discriminated significantly between children who were considered to be normal or deviant on a clinical rating (Richman, Stevenson, & Graham, 1982). Sensitivity was 90.2, while the specificity was 93.2. The 1-year stability of a deviant score was 63%, and a deviant score significantly predicted 62% of the children considered deviant on a clinical rating over the age range from 3 to 8 years. Cluster analysis of 183 normal and deviant children yielded five clusters mainly distinguishing between different toilet problems: Normal; Normal - Nightwet; Disturbed - Widespread sphincter control; Disturbed - Nightwet - Conduct; and Disturbed - Restless - Food fads.

Behaviour Checklist (BCL)

The BCL (Richman, 1977) for 3-year-olds has been modeled after the BSQ (Richman & Graham, 1971). Using 3- and 4-point scales, the same items of behavior as in the BSQ are completed by the parent quite independently, and summed into a total score. The checklist can be completed within 5 minutes. A factor structure has not been derived for the BCL. Richman (1977) reported a four-week test-retest reliability for the total score of $r=.81$. Using clinical ratings as criterion, scores above a cutpoint predicted almost 80 percent of the cases correctly. Using scores above the cutpoint, the stability of problems from age 3 to age 4 was 49%. The BCL has been reported to be used as a screening instrument to detect preschoolers at risk for behavior problems (Richman, Stevenson, & Graham, 1982).

Child Behavior Checklist for Ages 2-3 (CBCL/2-3)

The CBCL/2-3 (Achenbach, Edelbrock, & Howell, 1987) assesses in a standardized way the parents' view of their child's feelings and behaviors using 99 items, and one open-ended question to report problems not mentioned in the checklist. Respondents are requested to rate the behavior of the child on a 3-point scale, with *0* denoting that the item is not true of the child, *1* that the item is somewhat or sometimes true, and *2* that the item is very true or often true.

Achenbach (Achenbach, 1992; Achenbach et al., 1987) factor analyzed the CBCL/2-3 item scores in two samples of 2- to 3-year olds in 1986 (N=398) and in 1992 (N=546). The first sample included children from the general population, low birthweight children, and children referred to mental health services (24%). The second sample consisted of referred children (including those in the first sample), and children having a total problem score above the median in the American normative sample. In both samples, six components were obtained, and scales were composed of the items loading above a certain cutpoint, labeled Social Withdrawal (1986) or Withdrawn (1992), Depressed (1986) or Anxious/Depressed (1986), Sleep Problems, Somatic Problems, Aggressive (Behavior), and Destructive (Behavior). The latter four components were replicated fairly well across both analyses. As indicated by the change of labels, the components reflecting withdrawn and anxious behavior were less replicable (Achenbach, 1992). Further, two second-order factors, Internalizing and

Externalizing were found in both studies. Achenbach et al. (1987) and Achenbach (1992) reported a mean one-week test-retest reliability of $r = .87$ and $r = .85$, and a mean one-year stability of $r = .69$ and $r = .64$, respectively, for the scores on the CBCL/2-3. Children referred for mental health services scored significantly higher than nonreferred children on each syndrome and on the total problem score. A lack of significant correlations with the Minnesota Child Development Inventory (Ireton & Thwing, 1974), and with the Bayley (Bayley, 1969) and McCarthy (1972) scales of development indicated that the CBCL/2-3 taps problem behavior independently of the developmental variance tapped by these measures.

2.3.2 Teacher rating scales

Kohn Symptom Checklist

The Symptom Checklist (Kohn & Rosman 1972) was developed to cover the major clinical symptoms which young children manifest in preschool or in day care settings. It consists of 49 items to be filled out by teachers or teacher-aids on 3-step scales. Factor analyses yielded two (a priori defined) factors: Apathy-Withdrawal and Anger-Defiance, which showed stability over 6-, 12-, and 18-months intervals (mean $r = .69$, .49. and .41, respectively) for 3- and 4-year old children. Low to moderate cross-time relations have been found between Symptom Checklist factors and corresponding factors of the Peterson Problem Checklist. The Symptom Checklist shows moderate discriminative power between well functioning children and children in various treatment groups (Kohn & Rosman, 1973).

Preschool Behavior Questionnaire

The Preschool Behavior Questionnaire (PBQ; Behar & Stringfield, 1974) was originally developed to identify preschoolers at risk for the development of later emotional problems. It consists of 30 items using a 3-point rating scale. With preschoolers aged 3-6 years, Behar and Stringfield (1974) found three factors: Hostile-Aggressive; Anxious-Fearful; and Hyperactive-Distractable. The first two of these factors have been replicated in several studies (Fowler & Park, 1979; Hoge, Meginbir, Khan, & Weatherall, 1985; Tremblay, Desmarais-Gervais, Gagnon, & Charlebois, 1987). Test-retest reliability of the scale over a 3- to 4-month period ranged from .60 to .94, and interrater reliability for teacher ratings was .67 to .84 (Behar & Stringfield, 1974). The scale significantly differentiates between both hyperactive and emotionally disturbed and normal preschoolers (Behar, 1977; Campbell, Szumowski, Ewing, Gluck, & Breaux, 1982; Prior, Leonard, & Wood, 1983). PBQ ratings show significant correlations with classroom behavior and with scores on the Kohn Symptom Checklist (Behar, 1977; Behar & Stringfield, 1974; Rubin & Clark, 1983).

Preschool Behaviour Checklist (PBCL)

The PBCL (McGuire & Richman, 1986) was designed for screening emotional and behavioral problems in 2- to 5-year-olds in various preschool settings. The PBCL is a 22-item version of the BSQ for completion by preschool teachers. McGuire and Richman (1986) reported interrater agreement of 83%, using classifications above cutpoint, interrater score

correlations of $r=.60$, and internal consistencies of .83. The checklist discriminates significantly between psychiatric day centre and nursery samples. Factor analysis of the PBCL yielded three factors: Conduct-Restless-Aggressive; Isolated-Immature; and Emotional-Miserable (McGuire & Richman, 1986), which were clearly replicated in a Hong Kong preschool sample (Luk, Leung, Bacon-Shone, & Lieh-Mak, 1991). No psychometric data on these factors were reported.

2.3.3 Discussion of available rating scales

The teacher rating scales described above provide reliable and valid ratings of broad-band dimensions of preschoolers' problem behaviors. For preschoolers, only syndromes similar to the broad-band dimensions Internalizing and Externalizing have been replicated consistently. Factor-analyses of data obtained with the teacher ratings of problem behavior of preschoolers in the age range of 2 to 6 years, yielded an 'externalizing' dimension variously labeled Hostile-Aggressive, Anger-Defiance, and Conduct-Restless-Aggressive, and an 'internalizing' dimension labeled Anxious-Fearful, Apathy-Withdrawal, and Isolated-Immature (Behar & Stringfield, 1974; Kohn & Rosman, 1972; McGuire & Richman, 1986). Behar and Stringfield retained a third factor comprising four items with high loadings, that they labeled as Hyperactive-Distractable. However, this factor has not been replicated by others using the same instrument (Fowler & Park, 1979; Hoge et al., 1985; Tremblay et al., 1987). Thus, the analyses of teacher ratings leave open the question whether more differentiated dimensions of problem behaviors can be found among 2- and 3-year olds. Furthermore, although 3-year-olds were included in the study samples, the applicability of the dimensions to toddlers aged 2 years remains uncertain. The limitations inherent to the existing teacher ratings apply also to the most frequently used parent rating scale, the BSQ (Richman & Graham, 1971) and its derivative, the BCL (Richman, 1977). Although these appear to be useful screening instruments, they do not provide a well differentiated picture of particular problem areas, and no data on their validity for even younger children are available.

Achenbach and coworkers (Achenbach, 1992; Achenbach, Edelbrock, & Howell, 1987) were able to further differentiate between preschoolers' behavior problems. Using parent ratings on the Child Behavior Checklist for Ages 2-3, Achenbach et al. performed principal component analyses in two samples of 2-3-year-olds in 1986 and in 1992. In both samples six components were obtained, and scales were composed of the items loading above a certain cutpoint, labeled Social Withdrawal or Withdrawn, Depressed or Anxious/Depressed, Sleep Problems, Somatic Problems, Aggressive (Behavior), and Destructive (Behavior). Further, second-order groupings of Internalizing behavior, including the Withdrawn and Anxious/Depressed scales, and Externalizing behavior, including the Aggressive Behavior and Destructive Behavior scales were found. All scales discriminated well between nonreferred children and children referred for mental health services. Thus, the CBCL/2-3 seems a reliable and valid intrument to provide a more differentiated picture of young preschoolers' problems than may be obtained by other instruments. Therefore, it seems the

instrument of choice to further the study of psychopathology in preschoolers. Usage of the same standardized instrument across studies and cultures enhances the comparability of results from different studies. Before introducing the CBCL/2-3 in the Netherlands as a tool for research and diagnostic purposes, it is necessary to analyze to what extent it is applicable in this country. Three goals have to be attained to answer this question.

First, we need to establish the cross-national comparability of the CBCL/2-3 syndromes. Although comparisons of syndromes derived from the CBCL/4-16 (Achenbach & Edelbrock, 1983) for American and Dutch samples indicated high cross-national similarity for most syndromes (Achenbach, Verhulst, Baron, & Althaus, 1987; Verhulst, Achenbach, Althaus, & Akkerhuis, 1988), this is not necessarily the case with the CBCL/2-3. Sample differences as well as cultural differences may be responsible for differences in the CBCL/2-3 factor structure across studies. In performing principal components analyses on two different samples, Achenbach (1992; Achenbach et al., 1987) found high similarity for small-band syndromes reflecting externalizing problems, but considerably lower similarity for specific syndromes reflecting internalizing problems, i.e., Withdrawn and Anxious/Depressed. Furthermore, considering the second-order Internalizing and Externalizing factors, the position of the Withdrawn syndrome on the Internalizing factor seems not yet fully ascertained. In the 1986 second-order analyses (Achenbach et al., 1987) the two small-band syndromes Social Withdrawal and Depressed loaded .92 and .65, respectively, on the Internalizing factor. However, in the 1992 second-order factor-analyses on syndrome scales for both sexes (Achenbach, 1992), the mean loadings of the Withdrawn scale were of equal magnitude (0.50) on both second-order factors. Consequently, the observed correlation between Internalizing and Externalizing scores was high ($r = .75$).

Second, we need to assess the psychometric properties of CBCL/2-3 syndrome scales in a Dutch population: internal consistency, test-retest stability, and interrater agreement; construct and criterion related validity, and predictive validity.

Third, for diagnostic and clinical applications, we need to obtain Dutch normative data on the CBCL/2-3 items and syndromes from a general population sample.

The studies reported in this thesis were aimed at these three goals.

2.4 Prevalence of problem behaviors in preschoolers

To study the distribution of problem behaviors in the population of 2-3 year olds, it may be sufficient to obtain proportions of children in the population at large displaying specific problem behaviors by any means. To study the prevalence of significant disorder in a population one has to count the number of identified cases in a circumscribed population. However, we do not have unequivocal criteria to define psychiatric disorder in children. As pointed out elsewhere (Verhulst & Koot, 1992), different approaches to the definition of caseness have been employed in child psychiatric epidemiology: the assessment of individual problems; item counts; and clinical judgment.

Most *individual problems* involve behaviors that many children exhibit at some time during their development. These may only be regarded as pathological when very severe or

persistent, when they are concomitant with other problems, or when they have a negative impact on other domains of development. However, also rare behaviors, such as self-mutilation in a young child, or less rare behaviors which may show little association with overall dysfunction (such as bed-wetting) may be reason for clinical attention.

Item counts may involve the summation of scores obtained across all items of an assessment instrument that covers a broad range of clinically relevant items, or the summation of the scores of the items constituting a specific syndrome. These summed scores may be handled statistically either as a dimensional measure, ranking subjects from the least to the most disordered, or by employing cutoff scores as a categorical measure for classifying children into normal or disordered groups. The validity of the determination of a certain cutoff score, such as the 90th percentile of the cumulative frequency distribution of the problem scores obtained for normative samples of children, as the border between normal and pathological functioning may be tested against a morbidity criterion. In the absence of a so-called "gold standard" for defining psychopathology, either referral of a child for mental health service or clinical psychiatric judgment are most often used as a criterion of psychopathology.

As discussed above, *clinical judgment* uses the clinician's knowledge, experience, and skill to define a case. The child's current functioning is judged in relation to what is normal for his or her age, the degree of handicap and/or distress it causes to the child himself, the family or community, and the persistence of the behavior. The judgment may result in the formulation of the presence of clinically significant disorder in the clinician's own words, by using a standardized scale, or by the application of diagnostic systems such as DSM-III or ICD-10.

In the epidemiological studies of problem behavior in preschoolers that are presented in the next section, each of these approaches to the definition of caseness was used.

2.4.1 General population studies

Studies on the prevalence of problem behaviors of 2-year-olds are rare. Three-year-olds have been studied as a separate sample or have been included in a so-called preschool sample. The occurrence of problem behavior in preschool children has been studied with widely varying results.

The reported prevalence of significant disorder among preschoolers based on parental reports ranges from 6% to 50% (Cornely & Bromet, 1986; Earls & Richman, 1980; Jenkins, Bax, & Hart, 1980; Jenkins, Owen, Bax, & Hart, 1984; Larson, Pless, & Miettinen, 1988; Long, 1941; Macfarlane, Allen, & Honzik, 1954; Minde & Minde, 1977; Richman, Stevenson, & Graham, 1975). This variation in results may not be surprising, since prevalence rates were derived from diverse and sometimes unstandardized assessment procedures. Furthermore, a number of studies used unrepresentative samples and many studies differed in their criteria for deviance, i.e., simple percentages of reported behavior, and cutoff points based on either percentile scores or on ratings made by psychiatrists.

The most frequently used approach to study prevalence rates in preschool children has

been to obtain parent reports on a limited set of specific problem behaviors. By employing a cutoff point for the sum of problem scores above which children's behavior can be considered deviant, it is possible to determine the number of 'disordered' individuals. An example of this approach is the study by Richman, Stevenson, and Graham (1975) using the Behavior Screening Questionnaire (BSQ; Richman & Graham, 1971), a 12-item questionnaire used in interviews with parents to obtain scores for a number of problem behaviors. The authors estimated that at age 3, seven percent of the children from the general population in an outer London borough had marked problems, and 14 percent had mild problems. Using the same or a very similar instrument and cutoff points, Jenkins, Bax, and Hart (1980) reported that 11% of the 3-year-olds in North London had a moderate or severe problem, while Earls and Richman (1980) reported that 16% children aged 3 of West-Indian born parents living in London had a moderate or severe behavior problem. As assessed by the BSQ, prevalence rates of 24% (Earls, 1980) and 12% (Cornely & Bromet, 1986) moderate or severe behavior problems have been found for two populations in the United States, while Minde and Minde (1977) reported 6% of preschool children in a Toronto, Canada sample to have a moderate or severe behavior problem.

As discussed earlier, although the BSQ is reported to be a reliable and valid screening instrument to detect preschool children with problem behavior, it does not give detailed reports on children's problem behavior. Furthermore, the cutoff score on the BSQ that creates normal and abnormal ranges has been based on psychiatrists ratings of the severity of behavioral disorder from the same institute, which may only partially converge with ratings obtained from other observers.

Only two prevalence studies using the CBCL/2-3 have been reported to date. Larson, Pless, and Miettinen (1988) used the 1987 version of the instrument (Achenbach et al., 1987) in sample of 756 children aged 3 years in Montreal, Canada. The prevalence of one or more deviant scores on the syndromes of the CBCL/2-3, using cutoff points at the 98th percentile, was 11.1%. In this study it was also found that slightly more than half of the children, with one or more syndrome scores above the 98th percentile, also had a total problem score in the deviant range (above the 90th percentile). Achenbach (1992) reported 12% of nonreferred children having one or more deviant scores on the 1992 syndromes of the CBCL/2-3.

The meaning of results from these latter studies is unclear, mainly for two reasons. First, the prevalence estimates were based on selective samples. Larson et al.'s (1988) study secured the participation of only 34% of the target birth cohort. The prevalence figure derived from Achenbach's (1992) results is based on a sample of children from the general population who had been matched for sex, age, respondent, ethnicity, and SES with a sample of children referred to mental health agencies. Second, no information on the relation between deviant scores and other indices of problem behavior, like referral to mental health agencies or parental complaints, was reported in these studies.

In contrast to older age groups (see Verhulst, Berden, & Sanders-Woudstra, 1985) no rigorous prevalence estimates of psychopathology in children below age 4 are available for the Dutch population. Verhulst and Akkerhuis (1983) reported that 16% of a limited sample

of 100 3-year olds had a score of 10 or more on the BSQ. In the study reported in Chapter 6, we made an effort to obtain valid data on the prevalence of behavioral and emotional problems in the Dutch general population of 2- and 3-year olds. Prevalence estimates of disorders or problems should be based on highly representative samples from the general population. Further, completion rates as well as the quality of the assessment procedures determine whether the obtained data validly represent the entire population (cf. Achenbach, 1992). Using the CBCL/2-3 as a reliable and valid instrument of problem behavior, we obtained data in a fairly representative sample of children randomly selected from the Dutch province of Zuid-Holland. In addition, data on referral to mental health services, and need for professional help were obtained.

2.4.2 Clinical studies

There are only few published studies of psychiatric clinic populations of preschool children.

Rescorla (1986) studied the prevalence of four broad diagnostic categories in a sample of 204 boys and 70 girls aged 3-5 years seen for evaluation at the Yale Child Study center. Twenty-one percent of the children received a diagnosis severe atypical; 13% mild atypical; 44% reactive disorder, including children who would receive a variety of DSM-III diagnoses including oppositional, conduct, overanxious, attention deficit, and separation anxiety disorder; and 23% "other", mainly consisting of children with language delay, hyperkinetic syndrome, or minimal brain dysfunction. Eight factors based on 73 CBCL (Achenbach, 1978) items rated from clinical case records showed highly significant differences between diagnostic groups. A close interrelation between aggressive/destructive behavior and feelings of anxiety, sadness, and emotional instability was found in these preschoolers. More than one-third of the children in the sample were above the median on both an acting/out, externalizing factor, and an emotional, internalizing factor. Further, a strong association between aggressive/destructive, acting-out behavior and problems in impulse control and attention was found.

Beitchman, Wekerle, and Hood (1987) followed a sample of initially 129 children (82% boys) under age 5 referred to a psychiatric preschool program. Fourty-five percent of these children came from lower socio-economic status families, and 54% came from intact families. The following groups of DSM-III diagnoses were reported for this sample: conduct-type disorder (15%), attention deficit disorder (32%), emotional disorder (21%), developmental delay (21%). Ten percent received no diagnosis.

Lee (1987) reported on 129 preschoolers (60% boys) below age 6 (44% under age 4) evaluated at a military psychiatric clinic. Of these, 71% came from intact families. Fifty-eight percent of the presenting problems were of a conduct type (difficult to manage, temper tantrums, hyperactive, aggressive, stubborn, poor relationships with other children, and school-behavior problems). Clinical diagnoses included adjustment disorder (35%), disruptive behavior disorder (25%), disorder of elimination (15%), specific, global, or pervasive developmental disorder (41%), parent-child problems (10%), and other (10%).

Hooks, Mayes, and Volkmar (1988) studied a sample of 113 boys and 80 girls under age

5 evaluated at a child psychiatric clinic. Half of the sample were referred by pediatricians or child protective agencies. Less than 50 percent of the children lived with both parents, and two-third of the children were from lower-class families. In 86% moderate to severe stress was reported in the family. One third of the presenting problems regarded developmental delays, one third behavioral problems, while the remaining problems were equally divided between anxiety/affective disturbances and abuse/neglect. The presenting problem, however, did not differentiate between DSM-III diagnoses given by the clinician after evaluation. About ninety percent of all children with a final diagnosis of either emotional disorder or disruptive disorder or a parent-child problem presented with aggression or oppositionality. Even of children with pervasive developmental disorder, apart from 97% showing developmental delay, 52% had also aggression/oppositionality as a presenting problem.

Other studies of clinic referred children reported on very small or highly selective samples (e.g., Dahl, Cohen, & Provence, 1986; Sherman, Shapiro, & Glassman, 1983; Wolff, 1961).

Although these studies give some insight into typical presenting problems of preschoolers as well as diagnoses given after evaluation, they show several major flaws. None of these studies used standardized instruments to assess these problems. Likewise, although DSM-III criteria were applied in some of the studies, the studies used no comparable methods to aggregate diagnoses. Further, each study sample was derived from only one child psychiatric unit, and thus is very unlikely to be representative of the behavioral/emotional problems typical for this age group. The studies also did not differentiate between problems of children of age 3 years or younger and older preschoolers. No psychosocial correlates of presenting problems or diagnoses were reported in any of these studies. In our study we obtained data on a sample of 2-3-year-olds referred to 12 child guidance and mental health settings. The sample was sufficiently large to study different diagnostic groups. Furthermore, data on children's problems were obtained using standardized parents' reports as well as standardized codings from file charts. Finally, parents' reports and file information on problem behaviors were related to demographic and family background information on the preschoolers.

2.5 Correlates of problem behavior in preschoolers

One of the aims of child psychiatric epidemiology is the determination of the association of demographic and environmental variables with child psychopathological manifestations. These variables may include age, sex, race, socioeconomic status, urbanization, marital status, family composition, family functioning, parental psychopathology, and stressful life events. Cross-sectional and longitudinal studies using general population samples and commodity samples have yielded a number of factors that have been associated with the onset and/or persistence of problem behaviors in young children. Campbell (1990) gives an overview of studies and provides a summary list of relevant factors, which is in adapted form reproduced in Table 2.2.

Table 2.2
Factors associated with the development of behavior problems in preschoolers

Child characteristics	*Parenting skills*
Biological risk/vulnerability	Insensitivity/unresponsiveness
Age	Unavailability
Gender	Limited/negative affective involvement
Irritability/difficultness	Inappropriate developmental expectations
Uneven/delayed cognitive development	Overly harsh or lax control strategies
Deficits in social cognition/skills	
	Family environment & Social context
Family composition and integration	Low educational level of parents
	Unemployment/underemployment
One-parent family	Limited financial/material resources
Marital distress	Low social support
Parental personality problems	Inadequate institutional support
Physical and mental problems	Inadequate child care facilities
Interparental disagreement over child rearing	Family stresses outside nuclear family
Number of children	

Note. Table adapted from Campbell (1990).

Of the factors mentioned in this table it is not known whether they are causes, correlates, or reactions to a problem. For example, in a relatively comprehensive epidemiological study Richman et al. (1982) found that several family factors were related to behavior problems at age 3 and 8 years. Most notably, maternal feelings of depression and lack of emotional responsiveness of the mother showed cross-sectional and longitudinal relations with problem behavior. However, since information on maternal feelings and child behavior was obtained from the same source conclusions about causality are not warranted.

To obtain information on the relationship between potential risk factors and problem behaviors in Dutch preschoolers we investigated the relationship between information on family and child factors with problem behaviors as assessed by the CBCL/2-3. For most factors the relationship was studied in both the community and clinical sample. Information was obtained from interviews with one or both parents in the community sample and from clinical file information in the clinical sample. Part of the analyses focused on factual information, i.e., information that does not imply any judgment on the part of the informant or the interviewer.

Some studies have found that it is the number of risk factors rather than their nature that is the best determinant of outcome in older as well as in younger children (e.g., Rutter, Tizard, Yule, Graham, & Whitmore, 1976; Werner & Smith, 1982). This means that the more risk factors are present, the worse the outcome is, independent of the particular nature of the factors. Werner and Smith (1982) followed a sample of children from birth to adolescence. They identified moderate to severe perinatal stress, and a congenital defect at birth as biological risk variables. Risk factors in the caregiving environment were low level of maternal education, low standard of living, and a low rating of family stability. Child characteristics identified as a risk factor were infant activity level at year 1, and IQ. Children with severe early trauma frequently showed no later deficits unless the birth problems were combined with severe environmental circumstances. In this long-term study, the presence of four or more of these predictors at age 2 appeared to be a valid cutoff point for children at risk for serious learning and/or behavior problems by age 10 or 18. Furthermore, they identified factors that apparently protected children at risk at age 2 from developing problems at a later age: good temperament, favorable parental attitudes, low levels of family conflict, counseling and remedial assistance, small family size, and a small load of stressful life experiences. In the Rochester longitudinal study of children from 4 to 48 months, Sameroff, Seifer, Baroca, Zax, and Greenspan (1987) identified a set of risk factors that were predominantly found in lower SES groups, but affected child outcomes such as intelligence and social-emotional competence in all social classes. Risk factors were maternal mental health, maternal anxiety, parental control, maternal spontaneous interaction, education, occupation, minority status, family support, stressful life events, and family size. No single factor was determinant of outcome. The number of risk factors present in the family showed a linear relationship with outcome. Only in families with multiple risk factors was the child's competence endangered.

This same principle may apply in the present cross-sectional community study. In this study we similarly identified a set of correlates apart from the relationship between each single correlate and the level of problem behaviors at age 2 and 3.

2.6 Continuity and change of problem behavior in young preschoolers

Findings on stability

Large-scale longitudinal studies of problem behaviors in children and adolescents aged 4 years and older have shown medium to large stability coefficients of parent ratings of behavioral/emotional problems (Ghodsian, Fogelman, Lambert, & Tibbenham, 1980; Verhulst & Althaus, 1988; Verhulst, Koot, & Berden, 1990) and medium stabilities for teacher ratings (Verhulst & Van Der Ende, 1992). Using a categorical approach stabilities of deviancy have been found on parent reports of 31%-75%, depending on the definitions of deviancy (Graham & Rutter, 1973; Verhulst & Althaus, 1988; Koot & Verhulst, 1992), and stabilities as low as 9.4% using teacher ratings (McGee et al., 1985). Recently, McConaughy, Stanger, and Achenbach (1992) showed medium to large stabilities of parent ratings over a 3-year interval for comparable scales of the CBCL/4-18 (mean *r*s ranging from

.32 to .59). Predictive correlations between Time 1 parent ratings and Time 2 teacher ratings, however, were considerably weaker than parent-to-parent correlations (mean *r*s ranging from .05 to .31). Classification of children as deviant showed weaker predictive relations than did quantitative scores.

A variety of studies using laboratory-based measures, teacher rating scales, and parental reports suggest that also preschool problem behavior may show marked stability and may be predictive of later behavioral and emotional problems.

Campbell and others studied the stability of problem behavior in small samples of parent-identified hard-to-manage toddlers and preschoolers in a series of laboratory based studies (Campbell, Breaux, Ewing, & Szumowski, 1984; Campbell, Ewing, Breaux, & Szumowski, 1986; Campbell, Schleifer, Weiss, & Perlman, 1977; Campbell, Szumowski, Ewing, Gluck, & Breaux, 1982). In a three-year follow-up half of the original sample of children identified as having significant psychopathology at age 3 continued to display significant behavior problems, when reassessed at 6 years of age on structured laboratory tasks and maternal ratings (Campbell et al., 1986). Teacher ratings and classroom observations of the children supported the maternal and laboratory ratings. These findings are confirmed by the results of a follow-up study of pervasively hyperactive children identified in the general population (McGee, Partridge, Williams, & Silva, 1991), which showed that after 12 years only a quarter of the sample had "improved".

Moderate stabilities (r=.44-.56) were found for teacher ratings of problem behavior in children initially seen at ages 3 and 4 over periods ranging from one to five years in studies reported by Kohn and Rosman (Kohn, 1977; Kohn & Rosman, 1972, 1973).

Moderate to high stability of parent ratings of problem behavior has been found in community survey studies using the BSQ. Minde and Minde (1977) reported that BSQ scores at age 3 correlated .66 with scores obtained nine months later. Coleman, Wolkind, and Ashley (1977), using an adaptation of the instrument, reported stabilities of r=.55 between ages 3 and 4, r=.50 between ages 3 and 5, and r=.71 between ages 4 and 5. Using the BSQ in a large scale sample of 705 3-year-olds, Richman, Stevenson, and Graham (1982) followed-up 99 problem children and an equal number of randomly selected controls at ages 4 and 8 years. Sixty-three percent of the problem group, compared to 11% of children in the control group still had deviant BSQ scores at age 4 years. At 8 years these percentages were 43% for the problem group, and 20% for the control group on the Rutter questionnaire. At this age 62% of the problem group versus 22% of the control group received clinical ratings of mild, moderate or severe disturbance. Further, more children in the problem than in the control group (24% vs. 4%) had had contact with child guidance services because of a problem with the index child. A one-year stability of 55% in problems reported on the BSQ in a community sample of 691 preschoolers aged 3 years was reported by Cohen and Bromet (1992), and a 54% stability in a sample of 61 3-year-olds over a 3-year period by Garrison and Earls (1985), each using the CBCL as follow-up instrument.

At least two studies have reported on the stability and predictive validity of CBCL/2-3 scores. Achenbach (1992) reported that in a longitudinal study of 75 low birthweight and normal birthweight subjects over a 1 year period, stability correlations of the CBCL/2-3

American scale scores ranged from .50 to .78 with a mean *r* of .64. Destructive and Externalizing scales showed declines in scores, while Anxious/Depressed, Sleep Problems, and Internalizing scales showed increases. Longitudinal correlations in the same sample between CBCL/2-3 scores and corresponding CBCL/4-18 scale scores at ages 4-9 years were moderate to high at average. Average correlations from age 2 to later ages for the Internalizing, Externalizing, and total problem score were .40, .58, and .59, and from age 3 to later ages .53, .60, and .68. Problems reported at age 3 thus provided a better basis for prediction of future problem scores than did problems reported at age 2. The Aggressive Behavior, Externalizing and total problem scales of the parent-reported CBCL/2-3 correlated .45 to .48 with the teacher's reports on the TRF at age 6. Rose et al. (Rose, Rose, & Feldman, 1989; Rose, Feldman, Rose, Wallace, & McCarton, 1992) reported comparable findings on the CBCL/2-3 for a small sample of low SES children over a period from age 2 to 6.

It may be concluded that parent ratings of young preschool children's problem behavior show stabilities comparable to those obtained for older children. Information from the studies reviewed suggests that high levels of problem behaviors remain in about 60% of the children, and that externalizing behaviors are somewhat more predictive of later behavior problems than internalizing behaviors. There are some indications that in preschoolers the stability is higher among boys than girls (Richman, Stevenson, & Graham, 1982), and higher for externalizing than for internalizing problems (Fischer, Rolf, Hasazi, & Cummings, 1984; Kohn & Rosman, 1972). Finally, high initial scores, especially of an externalizing nature seem consistently related to persistence of problems in preschoolers (Campbell et al., 1986, 1987, 1990; Cohen & Bromet, 1992; Richman, Stevenson, & Graham, 1982).

Predictive correlates

Several large-scale epidemiologic and longitudinal studies on children and adolescents have identified family, environmental, and child variables that are concurrently and predictively related to behavioral/emotional problems in children and adolescents of age 4 and older (Anderson, Williams, McGee, & Silva, 1989; Berden, Althaus, & Verhulst, 1990; Bird, Gould, Yager, Staghezza, & Canino, 1989; Cohen, Brook, Cohen, Velez, & Garcia, 1990; Costello, 1989; Graham & Rutter, 1973; Graham, Rutter, & George, 1973; Holahan & Moos, 1987; Offord, Boyle, & Racine, 1989; Velez, Johnson, & Cohen, 1989). Significantly related family variables included family constellation, family discord, and family dysfunction. Environmental variables included SES, neighborhood characteristics, and stressful life-events. Significantly related child variables were early behavioral/emotional problems, temperament, and academic problems.

In the present longitudinal study on preschoolers we will focus on levels of problem behavior assessed at 2-3 years, family environment, temperament, language problems, parenting stress, and stressful life-events as correlates and predictors of problem behavior at Kindergarten.

There is consistent evidence that family environment variables such as those studied by Sameroff et al., are related to cognitive and socioemotional outcome in young children

(Rutter, 1989; Sameroff et al., 1987). These variables include low family occupational level, low maternal education, minority status, and lack of marital or family support (Campbell et al., 1986; Cohen & Bromet, 1992; McGee et al., 1984; Richman, Stevenson, & Graham, 1982). This relation seems to be fairly nonspecific, however. That is, the presence of a large number of these variables rather than the presence of each specific variable seems to pose children at risk for development and persistence of behavior problems.

Thomas and Chess (1977) emphasized the role that difficult temperament and parental ability to adjust to it play in the etiology of behavior problems. Several studies indicate a strong association between parent reports of difficult temperament and concurrent and later reports of behavior problems (Barron & Earls, 1984; Bates & Bayles, 1988; Bates, Maslin, & Frankel, 1985; Cameron, 1978; Campbell & Ewing, 1990; Goldberg, Corter, Lojkasek, & Minde, 1990; Guérin & Gottfried, 1986; Swets-Gronert, 1986; Wolkind & DeSalis, 1982). There is some, inconclusive evidence on the specificity of the predictive relation between temperament and later behavior problems. Difficult temperament, defined as frequent and intense expression of negative emotions, seems to be clearly predictive of externalizing problems and somewhat less of internalizing problems. Poor adaptability would be more predictive of internalizing problems (Bates & Bayles, 1988; Bates, Maslin, & Frankel, 1985; Guérin, & Gottfried, 1986; Swets-Gronert, 1986). In the relationship between temperament and behavior problems the quality of the mother-child relationship or interaction may play a mediating role. Several authors suggest that difficult temperament in infancy may lead parents to display either more dysfunctional controlling behavior (Crockenberg, 1987; Lee & Bates, 1985; Maziade et al., 1985, 1989) or more passivity (Van Den Boom, 1988), which may provoke oppositional or withdrawn behaviors in the child.

Other child variables that may be influential in the development of problem behavior during the preschool period include cognitive development, most notably speech and language delay (e.g., Stevenson, Richman, & Graham, 1985; McGee, Silva, & Williams, 1984), and perinatal risk. Although prematurity and birthweight have been found to be related to behavior problems (see Hoy, Bill, & Sykes, 1988, for a review), this influence may largely be through later cognitive development, parental and environmental stimulation, and temperament (Goldberg et al., 1990; Weisglas, Koot, Baerts, Fetter, & Sauer, 1993).

Data on the influence of the early mother-child relationship, especially quality of attachment, on problem behaviors up to 4 years later are somewhat inconsistent. Although some studies are affirmative (Erickson, Sroufe, & Egeland, 1985; Lewis, Feiring, McGuffog, & Jaskir, 1984) other are not (Bates & Bayles, 1988; Bates, Maslin, & Frankel, 1985; Goldberg et al., 1990). The reason for this may be that, although an anxious attachment relationship with the mother may pose children at risk for developing behavior problems, changes in family stress, maternal well-being, and maternal responsiveness may affect this hypothesized relationship (Erickson, Sroufe, & Egeland, 1985; Goldberg et al., 1990). Therefore, it may be of equal importance to study the effect of concurrent parental attitudes and responsiveness on problem behavior. A negative parenting attitude has been found to related to the stability of preschoolers' problem behavior both in large scale (Cohen & Bromet, 1992; Richman et al., 1982) and small-scale studies (Campbell & Ewing, 1990;

Egeland, Kalkoske, Gottesman, & Erickson, 1990). In several of these studies parenting attitudes and behaviors seemed to vary as a function of maternal depression. However, since parenting attitudes and behavior are immediately experienced by children, this variable may be of more direct influence than maternal depression, and even moderate effects of earlier depression (Zahn-Waxler et al., 1990).

Stressful life-events are included in the present study because they can have direct effects on problem behavior in preschoolers (Campbell & Ewing, 1990; Egeland et al., 1990) as well as indirect effects (via parenting; e.g., Crnic, Greenberg, Ragozin, Robinson, & Basham, 1983).

Some data on the stability of parental reports of 2-3-year olds' general problem behavior are now available. However, information on the predictive value of more differentiated syndromes identifiable at ages 2-3 years is still lacking. Furthermore, it is unclear to what extent high problem scores at young preschool age obtained by parent report are predictive of the parent's and teacher's perception of the presence of a problems at school entry that are important independent of behavioral/emotional problems.

In the present research we studied over a period of 2 years the predictive value of CBCL/2-3 problem scores in relation to reports of problem behavior by parents and teachers using the CBCL/4-18 and the TRF, and by reports of significant problems by both informants using questionnaires composed for this purpose. Further, we studied the contribution of CBCL/2-3 scores in predicting outcome independent of family environment variables, the child's temperament and language problems, the level of perceived parenting stress, and stressful life-events occurring during the interval between the two times of measurement.

CHAPTER 3

Samples and methods of data collection

The previous chapter resulted in the following questions with regard to the assessment, prevalence and correlates of problem behavior in toddlers:

1. What are the psychometric characteristics of the Dutch version of the CBCL/2-3?
2. What is the prevalence of problem behavior in Dutch toddlers aged 2-3 years in the community and in a broadly defined clinical sample as measured by the CBCL/2-3?
3. How are CBCL/2-3 scores in both samples related to demographic characteristics and characteristics of the psychosocial environment, and in the clinical sample also to clinically relevant variables, such as referral complaints and psychiatric diagnosis?
4. How predictive are CBCL/2-3 scores of behavioral/emotional problems at home and at school over a 2-year interval, as measured by parent and teacher ratings of problem behavior, and reports of significant problems?

3.1 Design of the study

The research reported in this thesis consists of three parts. The first part concerns the psychometric qualities of the CBCL/2-3, i.e., its factorial structure, reliability, construct validity, and criterion related validity. Therefore, this part is essentially instrumental (De Groot, 1961). The second part involves two cross-sectional epidemiological studies. These concern the prevalence of problem behaviors in children from the general population and in children referred to mental health settings, and factors associated with the occurrence of these behaviors. The goals of these studies was to provide data on the distribution of problem behaviors in both samples, to provide normative and baseline data for the CBCL/2-3 to be used in other studies, and to generate questions and hypotheses for further research (cf. Verhulst, 1985). The third part involves a prospective longitudinal study. The goal of this part was to study the predictive validity of the CBCL/2-3 regarding home and school problems at ages 4 and 5 years.

3.2 Samples

3.2.1 The community sample

The community sample was drawn in september 1989 from children in the Dutch province

of Zuid-Holland. This area encompasses over 3,200,000 people in environments ranging from rural to highly urbanized. The total population of 2-3-year old children with the Dutch nationality living in the province of Zuid-Holland was estimated to be about 83,000 in 1987 (CBS, 1989, personal communication). In composing the community sample only children having the Dutch nationality were selected. Foreign children were excluded from this sample, because language difficulties would create problems in obtaining reliable interview data. Furthermore, unknown cultural differences, especially differences in the parents' tolerance and judgment of their children's behavior, might be a factor complicating the comparison of data obtained for these children and those obtained for Dutch children. For example, Weisz et al. (1988) showed that compared with American parents Thai parents rated both internalizing and externalizing problems as less serious, less worrisome, less likely to reflect personality traits, and more likely to improve with time.

To obtain a random sample of 2-3-year-olds living in the province, the cooperation of the inoculation register of the province of Zuid-Holland was invoked. Parents of all newborn children who are registered in a municipal population register in the Netherlands receive a convocation for compulsory, but free periodical inoculation of the child against infectuous diseases at the local general health facility. Most, but not all toddlers receive those health services. The inoculation register includes about 95 percent of all children aged 2-3 years living in the province, except those living in Rotterdam. A sample of 400 children (199 boys, 201 girls) was drawn randomly and stratified by age and sex from the register. Additionally, 69 children (34 boys, 35 girls) were drawn, stratified by age and sex, from the Rotterdam municipal health service register, which includes all 2-3-year-olds, living in the municipality. The resulting sample included a total of 469 children (233 boys, 236 girls) from 109 sites spread over the province of Zuid-Holland. The children's birthdates were evenly distributed over the months between October 15, 1985 and October 14, 1987.

3.2.2 Outline of data collection in the community sample

Parents of eligible subjects were contacted between September 1989 and March 1990. A letter was sent to the parents of the 469 children explaining the purpose of the study, the way in which the child was selected, and indicating that an interviewer would contact them by telephone to make an appointment for an interview concerning "important parts of the daily life of young children, such as eating, sleeping, behavior with parents, somatic complaints, development, and possible problems" (see Appendix C.1, for the letter to the parents). The parents were contacted by telephone and visited by one of four trained female interviewers, who had an education at the master's level in special education or psychology. In all cases the mother was the prime respondent. In 78 cases (18.8%) the husband or partner of the respondent was also present, in 223 cases (53.6%) the target child or a sib, and in 9 cases (2.1%) another adult. The interviewer read the CBCL/2-3 problem items regarding the target child aloud, and scored the parent's responses. The interviewer was instructed only to score responses of the prime respondent. After completing the CBCL/2-3, the parent was asked questions about demographic characteristics of the child and his family, child care

arrangements, the child's health, mental health services received by the child and family members, the occurrence of stressful life events, and on the need for professional help for mental health problems of the child (see Parent Interview I). The duration of the interview was 30-60 minutes. After the interview, the Behaviour Checklist (BCL) and the Minnesota Child Development Inventory (MCDI) were explained and handed over to be filled out by the parent, together with a postpaid envelope to return these questionnaires. In four cases the parent could not be interviewed, because the parent refused to be visited, or because the visit was interrupted. In these cases, the ratings scales were mailed to the respondent or left behind by the interviewer. The CBCL/2-3, Parent Interview, BCL, and MCDI were completed by 420, 416, 394, and 393 of the parents, respectively. After the visit parents were promised to be sent a concise general report of the study, and a concise report concerning the development of the individual child based on the data obtained with the MCDI. These were sent to the parents one month after the end of the data collection.

3.2.3 Demographic characteristics of the community sample

Of the 469 parents in the sample, 6 could not be reached, i.e., the parents were moved without notice of their new address to the municipal register or neighbours, 2 had moved to another country, 2 could not be interviewed because of language problems, and 39 refused to participate for various reasons. For 420 children (215 boys, 205 girls) interviews with the parent were completed, and could be used for data analysis. The response rate, corrected for 8 untraceables and 2 who could nor be interviewed, was 91.5% (89.5% uncorrected). Table 3.1 shows the demographic characteristics of the community sample. In 98.3% of the cases the respondent was the mother, in 1.4% the father, and in 0.2% both parents. Ninety-five percent of the cases were Dutch, 1.6% were Surinam, 0.2% came from the Dutch Antilles, 0.2% were Turkish, and 3.1% had another nationality. Parental educational level was coded according to a 9-step scale (ITS; Van Westerlaak, Kropman, & Collaris, 1975), which was recoded for purpose of analysis to a 4-step scale (cf. Standaard Onderwijs Indeling, CBS, 1987). The mean educational level of mothers was 2.56 (SD=0.80), and of fathers 2.74 (SD=0.87). Parental occupational level was scored on a 6-step scale (ITS; Van Westerlaak, Kropman, & Collaris, 1975). The mean occupational level of mothers was 3.53 (SD=1.41), and of fathers 3.68 (SD=1.44). The mean maternal age was 31.5 years (SD=4.4), and the mean paternal age was 34.2 years (SD=5.0). Of the 420 children, 10 (2.4%) had been referred to a child mental health agency within the past 12 months. For statistical analyses involving the comparison of the community sample and the clinical sample, a normative sample was composed by excluding the 10 referred children from the community sample.

Table 3.1
Demographic characteristics of the community sample and the clinical sample

	Community sample		Clinical sample	
	n	%	*n*	%
Child characteristics				
Total *N*	420	100	458	100
Sex				
Boys	215	51.2	302	65.9
Girls	205	48.8	156	34.1
Age				
Boys				
2;0 - 2;6 years	46	11.0	73	15.9
2;7 - 2;12 years	54	12.9	72	15.7
3;0 - 3;6 years	55	13.1	73	15.9
3;7 - 3;12 years	60	14.3	84	18.3
Girls				
2;0 - 2;6 years	51	12.1	40	8.9
2;7 - 2;12 years	52	12.4	29	6.3
3;0 - 3;6 years	51	12.1	39	8.5
3;7 - 3;12 years	51	12.1	48	10.5
Ethnic group				
Dutch	399	95.0	362	79.9
Surinam/Antillean	7	1.6	29	6.4
Mediterranean	1	0.2	9	2.0
Other	13	3.1	15	3.3
Dutch/Surinam	0	0.0	8	1.8
Dutch/Mediterranean	0	0.0	5	1.1
Dutch/Other	0	0.0	25	5.5
Unknown	0	-.-	5	-.-
Place of residence				
Both biological parents	385	92.5	341	74.6
Biological parent and partner	10	2.4	19	4.2
Biological parent	21	5.0	85	18.6
Adoptive family	0	0.0	5	1.1
Relatives	0	0.0	3	0.7
Foster parents	0	0.0	3	0.7
Residential setting	0	0.0	1	0.2
Unknown	4	-.-	1	-.-

Table 3.1 (Continued).

	Community sample		Clinical sample	
	n	%	*n*	%
Family characteristics				
Presence of mother figure				
Biological mother	416	100	442	97.1
Foster/adoption mother	0	0.0	8	1.8
Stepmother	0	0.0	1	0.2
No motherfigure	0	0.0	4	0.9
Unknown	4	-.-	3	-.-
Presence of father figure				
Biological father	394	94.7	360	79.8
Foster/adoption father	0	0.0	7	1.5
Stepfather	1	0.2	16	3.5
No fatherfigure	21	5.0	68	15.1
Unknown	4	-.-	7	-.-
Relationship between parent figures				
None/widowed/divorced	21	5.0	81	18.0
Married	385	92.5	323	71.6
Separated/LAT-relation/coparents	0	0.0	21	5.3
Cohabiting	10	2.4	26	5.8
Unknown	4	-.-	7	-.-
Number of children in the home				
1	92	22.2	125	27.6
2	218	52.5	242	53.4
3	79	19.0	61	13.5
≥4	26	6.2	25	5.5
Unknown	1	-.-	5	-.-
Position among siblings				
Firstborn	179	43.0	232	51.2
Second born	165	39.7	160	35.3
Third born	53	12.7	45	9.9
Fourth born or higher	19	4.6	16	3.5
Unknown	4	-.-	5	-.-

Table 3.1 (Continued).

	Community sample		Clinical sample	
	n	%	*n*	%
Parental educational level				
Mother				
Elementary school	16	3.8	42	11.5
Lower level secundary or professional	219	52.1	197	54.3
Medium level secundary or professional	118	28.1	80	22.0
University or high level professional	67	16.0	44	12.1
Unknown	0	-.-	95	-.-
Father				
Elementary school	15	3.6	34	10.5
Lower level secundary or professional	177	43.0	151	46.5
Medium level secundary or professional	120	29.1	66	20.3
University or high level professional	100	24.3	44	22.8
Unknown	8	-.-	133	-.-
Parental occupational level				
Mother				
Unskilled employees	12	2.9	22	4.9
Skilled manual employees	14	3.3	25	5.6
Clerical, technicians, minor professionals	55	13.1	52	11.6
Owners of small businesses	4	1.0	1	0.2
Supervisory, lesser professionals	40	9.5	19	4.2
Executives, major professionals, owners of large businesses	8	1.9	8	1.8
Housekeeper	277	66.0	316	70.4
Unemployed	10	2.3	6	1.3
Unknown	0	-.-	9	-.-
Parental occupational level				
Father				
Unskilled employees	8	1.9	62	15.4
Skilled manual employees	85	20.6	111	27.6
Clerical, technicians, minor professionals	124	30.1	81	20.1
Owners of small businesses	40	9.7	26	6.5
Supervisory, lesser professionals	76	18.4	42	10.4
Executives, major professionals, owners of large businesses	59	14.3	42	10.4
Housekeeper	1	0.2	10	2.5
Unemployed	19	4.6	28	7.0
Unknown	8	-.-	56	-.-

Table 3.1 (Continued).

	Community sample		Clinical sample	
	n	%	*n*	%
Income				
Mother				
No income	285	68.5	270	68.3
Job	115	27.6	80	20.3
Social assistance, pension, etc.	16	3.8	45	11.4
Unknown	4	-.-	63	-.-
Father				
No income	7	1.7	4	1.1
Job	383	92.0	325	88.8
Social assistance, pension, etc.	7	1.7	37	10.1
Unknown	23	-.-	92	-.-

Table 3.2
Distribution[a] of urbanization for the total population of the province of Zuid-Holland and for the community sample

Urbanization	Zuid-Holland (N = 3,200,408)	Community sample (N = 469)	Refusers (N = 39)
1. Rural; more than 20% of the population has agricultural profession	8.5	13.3	2.6
2. Semi-rural, less than 20% of the population has agricultural profession, fewer than 30,000 inhabitants	9.0	11.8	7.9
3. Suburban, less than 20% of the population has agricultural profession, more than 30% commuters	25.6	27.6	23.7
4. Small town with historical character	1.4	1.9	0.0
5. Urban	55.5	45.5	65.8

[a] On January 1, 1989 (Source: CBS, 1989)

Because the sample was not stratified by urbanization, we computed post hoc the urbanization grade of the areas in which the children lived. Table 3.2 shows the distribution of urbanization for the community sample and for the province of Zuid-Holland. The sample

shows a 10% underrepresentation of the urban areas, and a 2.0 to 4.8% overrepresentation of the rural, semi-rural and suburban areas. Refusers, however, were significantly overrepresented in urban areas and underrepresented in rural areas ($\chi^2=8.88$, $df=3$, $p<.05$).

There is evidence that refusers have somewhat higher problem scores than responders, at least in school-age and adolescence (Verhulst, Koot, & Berden, 1990). Therefore, we tested differences in CBCL/2-3 total problem scores (see further) between the urbanization groups in a 4 (urbanization) x 2 (age) x 2 (sex) ANCOVA design with SES as covariate. Children from suburban areas and boys from semi-rural areas had somewhat higher problem scores than children from urban and rural areas and than girls from semi-rural areas ($F(3, 394)=3.53$, $p<.001$), although the proportion of variance accounted for by urbanization was only 2.8%.

Two years after the first time of measurement, the community sample was approached again in a follow-up study. The two times of measurement will be designated as Time 1 and Time 2, respectively. The data collection procedure at Time 2, and characteristics of the sample participating at Time 2 will be described in Chapter 8.

3.2.4 The clinical sample

The clinical sample consisted of 458 children (302 boys, 156 girls) referred to a total of 12 child guidance and mental health settings (listed in Appendix C.2): four child psychiatric outpatient clinics, three of which were university clinics and one operated by an independent hospital; four outpatient community mental health centers; three day clinics for children up to age six, who are maladjusted or who show developmental problems; and one child guidance consultation agency. Three agencies operated outside Zuid-Holland: two of the university clinics operated in urban areas, and one day clinic operated in a rural area. All other agencies operated in urban or suburban environments in Zuid-Holland. Children were referred for behavioral and emotional problems and developmental delays. The data were collected between September 1987 and August 1991.

3.2.5 Outline of data collection in the clinical sample

The settings were asked to have parents or others in custody who came with the child fill in the CBCL/2-3 as part of their intake procedure (see Appendix C.3, for the introductory letter to the parents). In preparation of the data collection, mental health workers and office personnel, who would be involved in any way, were instructed on the purpose and procedures of the study. Comparable to the instructions given to the interviewers of the community sample, the personnel was informed how to answer the parents' questions, so that the conditions under which the checklists were filled out were approximately the same. CBCL/2-3's were handed over to the parents and caretakers at intake, filled out by them at the office or at home, and checked by the mental health worker. If possible, language difficulties were solved with the help of ethnic mental health or social workers. Responders

were informed that the checklist would be used for purposes of guidance and treatment as well as for research. To assist services with the use of the checklist for clinical purposes, completed checklists were entered and scored using the 1987 CBCL/2-3 computer program at the treatment setting, or sent to the investigator, who entered and scored the checklists, and sent back the checklists and scored behavior profiles. A completed checklist was included in the study, only if no items were missing and if we were able to obtain clinical file information on the child as well. When two or more checklists were completed for a child at intake, only the checklist was retained for analysis that was completed by the informant who most likely could be regarded as the daily caretaker of the child, because she or he ran the household. For children, for whom more than one checklist was completed, because they were referred to more than one of the participating agencies in the period of data collection, only the checklist completed for the first occasion was retained for analysis. When two checklists were filled out by each parent individually and independently on the same occasion, these checklists were also used in our analysis of interparent agreement (see Section 4.3).

3.2.6 Demographic characteristics of the clinical sample

Table 3.1 shows the demographic characteristics of the clinical sample. In this sample, 66.8% of the respondents were mothers, 6.2% were fathers, and 27.4% were both parents or others who were in custody of the child. The child's ethnicity was based on information of the origin of both parents, if available. If a child had only one Dutch parent, the child was designated to the ethnic group of the other parent. According to this classification, 79.9% of the referred children were Dutch, 8.2% were Surinam/Antillean, 3.1% came from the Mediterranean countries, and 8.8% came from other ethnic groups. Parental educational and occupational level were coded according the same scales as used in the community sample. The mean educational level of mothers was 2.35 (SD=.84), and of fathers 2.55 (SD=.96). The mean occupational level of mothers was 2.95 (SD=1.44), and of fathers 3.00 (SD=1.60). The mean maternal age was 30.3 years (SD=5.0), and the mean paternal age was 33.9 years (SD=5.9). Fifteen of the children (3.3%) were part of a twin pair.

3.3 Measures

3.3.1 Measures used in both the community and clinical sample

Child Behavior Checklist for Ages 2-3 (CBCL/2-3)

To obtain quantitative scores of problem behavior in both samples we used the CBCL/2-3 (Achenbach, 1992; Achenbach, Edelbrock, & Howell, 1987; see Appendix C.4). The CBCL/2-3 was translated into Dutch by F.C. Verhulst. It contains 2 pages to be filled out by parents or caretakers of children aged 2 and 3 years. It has been modeled on the Child Behavior Checklist for Ages 4-16 (CBCL/4-16; Achenbach & Edelbrock, 1983), and consists of 99 items on behavioral and emotional problems, and one additional item for problems that

have not been mentioned in the checklist. Fifty-nine of the items have counterparts on the CBCL/4-16, while the remaining items have been developed specifically for ages 2-3. Examples of items are *Stubborn, sullen, or irritable*; *Looks unhappy without good reason*; *Physically attacks people*; *Too fearful or anxious*; and *Has trouble getting to sleep*. Details on the development of the CBCL/2-3, and on the reliability and validity of the American version are given by Achenbach et al. (1987), by Achenbach (1992), and in Chapter 2.

The CBCL/2-3 requires only fifth-grade reading skills to complete. Most respondents can complete the form in less than 10 minutes. Respondents are requested to rate the items that describe the child *now or within the past 2 months* as *2* if the item is *very true or often true* of the child, as *1* if the item is *somewhat or sometimes true* of the child, and as *0* if the item is *not true* of the child. On 12 items the respondent is asked to describe the behavior, making it possible to correct the scoring when necessary, and to prevent more than one item from being scored for the same problem. Achenbach et al. (1987; 1992) constructed six syndrome scales, i.e., Anxious/Depressed, Withdrawn, Sleep Problems, Somatic Problems, Aggressive Behavior, and Destructive Behavior by principal components analysis of the 99 items. A second-order factor analysis, i.e., a factor analysis of the six syndromes, yielded two groupings of syndromes, Internalizing and Externalizing. Syndrome scale scores and Internalizing and Externalizing scores can be computed by summing the scores (0, 1, or 2) of the items comprised in these syndromes. A total problem score is obtained by summing the 0's, 1's, and 2's rated for each item, and has a theoretical range from 0 to 200. Our study on the factor structure and the psychometric characteristics of the Dutch version of the CBCL/2-3 will be reported in Chapters 4 and 5.

3.3.2 Measures used only in the community sample

To test the discriminant construct validity of the CBCL/2-3, i.e., to what extent it correlates more highly with a measure of the same or a related construct than with a measure of an unrelated construct, the Behaviour Checklist (BCL; Richman, 1977) and the Minnesota Child Development Inventory (MCDI; Ireton & Thwing, 1974) were used.

Behaviour Checklist (BCL)

The BCL (Richman, 1977) for 3-year-olds is modeled after the Behaviour Screening Questionnaire (Richman & Graham, 1971), which is completed by an interviewer during an interview with the child's parents. The BCL uses the same items of behavior as the BSQ, but in contrast to this it is filled in by the parent quite independently. The BCL consists of 19 items regarding 12 problem areas: eating, soiling, sleeping, activity, concentration, dependency/attention seeking, control, tempers, mood, worries, fears, and behavior with siblings and peers. The parent is asked to choose which behavioral description out of 3 or 4 choices best fits their child over the past 4 weeks. The checklist can be completed within 5 minutes. Although no official scoring system has been published, the authors of the checklist summed the highest ratings (0, 1, or 2) in each problem area covered by one or more items to produce the BCL score (P.J. Graham, personal communication). We used the

same procedure to obtain a total score.

The BCL has been reported to be used as a screening instrument to detect preschoolers at risk for behavior problems (Richman, Stevenson, & Graham, 1982). Richman (1977) reported a four-week test-retest reliability of $r=.81$. Using a score of 10 points or more, the sensitivity of the BCL was 69.6 and the specificity 87.4 with a clinical rating of mild, moderate, or severe behavior problems as criterion (21.5% misclassifications), which compares to a sensitivity of 90.2 and and a specificity of 93.2 for the BSQ (8.3% misclassifications). The 1-year stability of behavior problems from age 3 to age 4 as indicated by a score above the cutoff point on the BCL is 49%, as compared to 63% using the BSQ. Notwithstanding the somewhat lower psychometric quality of the BCL as compared with the BSQ, apart from the CBCL/2-3 it is the only currently available parent rating of problem behavior for this age group that can be filled out independently. Therefore we used this checklist as a measure for the discriminant construct validity of the CBCL/2-3.

Minnesota Child Development Inventory (MCDI)

The Minnesota Child Development Inventory for 0- to 6-year-olds (MCDI; Ireton & Thwing, 1974) asks a true/false response of parents on 320 empirically derived items describing children's development and behavior. The caregiver is asked to respond "yes" or "no" to all of the items, which takes approximately 45 minutes. The items form eight developmental scales: *Gross Motor*, *Fine Motor*, *Expressive Language*, *Conceptual Comprehension*, *Situation Comprehension*, *Self-Help*, and *Personal-Social*, and a *General Development* scale, which is composed of the 131 items from the seven other scales best discriminating between age groups. Because the MCDI is relatively unknown in the Netherlands as a measure of developmental status, we will give a short review of reported validity studies.

Several studies have demonstrated the criterion-related validity of the MCDI. Ireton, Thwing, and Curier (1977) reported that specific MCDI scales corresponded highly with developmental evaluations of a group of clinically referred children of which a substantial percentage was mentally retarded. Subsequent researchers found that the MCDI effectively discriminated between behavior problem and nonproblem preschoolers (Garrity & Servos, 1978), and between Head Start and nursery school children (Ullman & Kausch, 1979). The MCDI has been shown to correlate significantly with objective developmental measures. Colligan (1976, 1977), and Eisert, Spector, Shankaran, Faigenbaum, and Szego (1980) reported significant correlations between MCDI scales and developmental tests in high risk children. Using a short form of the McCarthy Scales of Children's Abilities (MSCA), Dean and Steffen (1984) found a significant correlation between subtest composites and the MCDI. Gottfried, Guerin, Spencer, and Meyer (1983, 1984) found that the MCDI General Development and language scales correlated positively and significantly with all cognitive scales of the MSCA in a sample of middle-class preschool children, which was confirmed by others also for pediatric samples (Byrne, Backman, & Smith, 1986; Kenny, Hebel, Sexton, & Fox, 1987; Byrne, Smith & Backman, 1987). Gottfried et al. (1984) also showed screening efficiency of the MCDI as well as its short-term predictive value across 6-12-month

intervals. Positive and high correlations were found between Bayley Mental Scale scores and MCDI scales for infants up to 30 months (Saylor & Brandt, 1986). Further, the language scales have a moderate to strong correlation with measures of expressive language (Tomblin, Shonrock, & Hardy, 1989). Guerin and Gottfried (1987) studied the cross-time consistency and the predictive validity of mother's reports on the MCDI from 2 1/2 to 5 and 6 years. The MCDI had a predictive value and structure comparable to that of objective tests (such as the MSCA and the Kaufman Assessment Battery for Children) with regard to intellectual and academic performance and social adaptive functioning at the school entry years.

We thus may conclude that the MCDI, particularly the General Development and language scales, provides a valid indication of the preschooler's developmental status. The reason for including the MCDI in this study was to use it as a measure for the discriminant construct validity of the CBCL/2-3.

We translated the MCDI into Dutch. To construct sex- and age-related norms for the MCDI, for each sex eight three-months age groups were formed, and for each group frequency distributions of the scale scores were computed. Using these distributions, the raw scores for each child were substituted by sex- and age-appropriate percentile scores. These percentile scores were used in the analyses.

Parent Interview I (PI-1)

The Parent Interview (PI-1) was designed as a semi-structured interview to obtain information on demographic variables and on family social background and risk. At Time 2, a second Parent Interview (PI-2) was conducted to obtain information on family constellation, health, behavioral outcome, the need of professional help, and the use of services (see Parent Interview II). This interview also included items on birth problems, that were not in the Time 1 interview, but which were clearly applicable to the Time 1 situation. Information from both interviews applicable to the first time of measurement was used to construct environmental, family, and child variables to be related to the child's problem behavior scores. The construction of these indices will be described in Chapter 6.

Life Events Questionnaire (LEQ)

The Life Events Questionnaire (LEQ; Berden, Althaus, & Verhulst, 1990) was designed to obtain information on life events that may be regarded as stressors, and therefore constitute a possible risk factor for the development of behavior problems. The at that time still preliminary version of the LEQ contained 34 items that had sufficient reliability. For example, its test-retest agreement was 74% and 83%, depending on whether the data were obtained by the same or by a different interviewer. The parent was asked to indicate whether the event had happened at any time since the birth of the child, the date of the event, and to rate whether the event implied a difficult period for the child or the family or both, using a 3-step scale (0=*No*; 1=*Somewhat*; 2=*Definitely*). Results of theoretical and empirical analyses by Berden (1992) guided the composition of the score based on the life-events questionnaire. First, the analyses revealed that only unambiguously negative life-events were related to an increase in problem behavior in children 4-16 years of age. Therefore, items

that were not stated as or implying a negative event were removed from the questionnaire. We removed the following items: *1. Birth of sib*; *6. Child went to nursery school*; *9. Severe illness of friend*; *11. Moved to another home*; *12. Child lost contact with friend because of removal*; *23. Parent disabled to work*; *27. Caregiving arrangement other than with parents*; *28. Severe illness of grandparent*; and *30. Adult moved into the family*. Second, items that imply problem behavior themselves may be spuriously related with problem behavior scores. Therefore, we also removed item *7. Increase in conflict between child and parents*. Third, events that happened more than one year ago are not reliably reported by parents. Thus, we only counted events that happened within the period one year before the interview. Fourth, parental ratings of the impact of the event on the child and the family were not stronger related to problem behavior than the sheer number of events. As a result, our life events score was a simple summation of the 24 remaining items scored as present in the year before the interview.

At Time 2, a 32-item questionnaire based on the work of Berden (1992) was used to assess the number of life-events, that had occurred between Time 1 and Time 2 of the study (see Chapter 8).

3.3.3 Measures used only in the clinical sample

File Information Form (FIF)

To obtain background information on the clinical sample, we constructed a coding scheme designated as the File Information Form (FIF). The FIF was used to code anonymously demographic and diagnostic information from the individual case files available at the participating agencies. It has 40 items that cover demographic characteristics of the child (6 items), demographic characteristics of the family (16 items), information on referral and obtained services (5 items), and diagnostic information obtained during intake and initial assessment (13 items). Following Veerman (1990), variables were chosen that could preferably be coded objectively, and for which sufficient information could be expected to have been obtained at all agencies as part of the routine intake procedure and initial assessment, and to be found in the child's personal file at the agency. All items regarded the situation at, or immediately preceding intake, except birth problems, hospital stays, and stability and quality of caregiving environment, which regarded the period from birth to intake. In addition, the coding of ICD-10 Axis V involved the 12 months preceding intake. A manual containing coding categories and coding rules for each item guided the coding of the information from the file onto the FIF. Nine items asked for a rating of the available

Table 3.3
Reliabilities of the File Information Form coding scheme

	Intrarater Agreement		Interrater Agreement	
	%	*K*	%	*K*
Demographic characteristics				
Birth date	100	-	100	-
Sex	100	1.00	100	1.00
Age mother	96.7	.93	90.0	.84
Age father	96.7	.92	83.3	.79
Marital status mother	93.3	.80	86.7	.56
Marital status father	90.0	.76	86.7	.66
Relationship between parents	90.0	.65	83.3	.55
Presence of mother figure	93.3	.55	93.3	.47
Presence of father figure	86.7	.42	83.3	.21
Place of residence	100	1.00	93.3	.80
Number of children in the home	100	1.00	100	1.00
Position among siblings	96.7	.93	90.0	.83
One of twin pair	100	1.00	100	1.00
Ethnic background mother	100	1.00	93.3	.64
Ethnic background father	90.0	.65	86.7	.29
Educational level mother	90.0	.75	46.7	.25
Educational level father	80.0	.64	60.0	.43
Occupational level mother	86.7	.71	83.3	.69
Occupational level father	93.3	.73	80.0	.64
Job mother	73.3	.58	66.7	.51
Job father	93.3	.77	90.0	.73
Income mother	80.0	.62	66.7	.50
Income father	93.3	.76	90.0	.73
Referral and diagnostic information				
Referral source	80.0	.67	73.3	.62
Referral question	90.0	.78	80.0	.60
Number of services obtained in the past	56.7	.31	60.0	.37
Planned services	83.3	.64	60.0	.51
Placement in (semi)residential care	90.0	.68	86.7	.65
Complaints at intake	83.7	.78	73.3	.62
Birth problems	93.3	.85	83.3	.66
Hospital stays	90.0	.74	63.3	.53
Mental problems mother	76.7	.54	66.7	.46
Mental problems father	83.3	.41	80.0	.31
Stability of caregiving environment	73.3	.54	50.0	.22
Quality of caregiving environment	73.3	.61	56.7	.28
ICD-10 Axis I	83.3	.71	70.0	.62
ICD-10 Axis II	66.7	.59	53.3	.37
ICD-10 Axis III	73.3	.59	66.7	.51
ICD-10 Axis IV	86.7	.78	73.3	.65

information: mental problems of mother, mental problems of father, stability of caregiving environment, quality of caregiving environment, and the five axes of the ICD-10. To code stability and quality of caregiving environment, coding rules provided by Veerman and Tates (1987) were used. ICD-10 Axes I through IV were coded according to the 1989 draft of ICD-10 Chapter V (Mental and behavioral disorders; WHO, 1989), and ICD-10 Axis V according to the draft ICD-10 Axis Five (Associated abnormal psychosocial conditions; WHO, 1988).

The coding of files was done by two raters who had an education at the master's level in special education (95 and 88 files, respectively) and by the author (HK; 275 files). To assess the reliability of the coding, after 6 months the first rater and the author each coded 15 files again, and the author coded 30 files also coded by the first rater. All files came from the same agency, because this was the only one to which both raters had access. From these codings both the intrarater reliability and the interrater reliability could be computed for 30 cases. For each item percentages of agreement and Cohen's Kappa (1960), to correct for chance agreement, are given in Table 3.3. These have been ordered by demographic and family correlates, and referral/diagnostic categories. Kappas were computed using the formula $K=(Po-Pc)/(1-Pc)$, where Po denotes the proportion of observed agreement, and Pc denotes the proportion of agreement expected by chance. Reliabilities could not be computed for the second rater because she left for another job before the reliability procedure started.

Computation of reliability coefficients for ICD-10 Axis Five was considered inappropriate because the overall proportion of missing values was .47. Landis and Koch (1977) provide rules for the interpretation of Kappa: Kappas of .21-.40 are considered fair, of .40-.60 moderate, of .61-.80 substantial, and of .81-1.00 almost perfect. The *intra*rater reliability of both the demographic and referral/diagnostic variables was substantial. Of the 38 Kappas computed on intrarater agreement nine (23.6%) were below .61. Overall, the *inter*rater reliability was lower. Of the demographic variables 40.9% were below .61; of the referral/diagnostic variables 56.2% were below .61. Although most variables may be considered to have adequate reliability for research purposes (Herjanic & Reich, 1982), results obtained using some variables showing only moderate intra- and interrater reliabilities should be interpreted with caution. These were: presence of motherfigure and fatherfigure; job of mother; number of services obtained in the past; mental problems of both parents; stability of caregiving environment; ICD-10 Axes II and III.

Low interrrater reliablities for diagnostic information were also found by Veerman (1990), using similar categories. The moderate interrater reliabilities found in the present study may partly be attributed to different use of the category "unknown" by the raters. The mean proportion of diagreement due to the differential use of the category "unknown" was .30.

3.3.4 Measures used in the follow-up of the community sample

Problems at home

Child Behavior Checklist for Ages 4-18 (CBCL/4-18)

The Child Behavior Checklist (CBCL/4-18; Achenbach, 1991b) was used to obtain standardized parents' reports of the behavioral/emotional problems and competencies of the children. The CBCL/4-18 is part of a "family" of highly similar instruments based on the CBCL to obtain reports of behavioral/emotional problems from parents, teachers (Teacher's Report Form; Achenbach, 1991c), children themselves (Youth Self-Report; Achenbach, 1991d), and young adults (Young Adult Self-Report). It consists of 20 competence items and 118 items concerning behavioral/emotional problems. The competence items ask to specify the number of sports, hobbies, organizations, jobs, and friendships the child takes part in as well as the amount and quality of participation. Other items ask to specify how well the child gets along with siblings, peers, and parents, and how well the child plays and works alone. Finally, items ask about the child's performance in academic subjects and school related problems. The response format is different for most of the items. The 118 problem items describe a broad range of problems that are of concern to parents and clinicians. Several items require parents to describe the problem briefly. This makes it possible to correct the parents' scoring when the description does not fit the item. Two open-ended items make it possible for parents to add other physical problems and any other problems not specified elsewhere. Parents are requested 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 it is *very true or often true*. The CBCL/4-18 is self-explanatory, enabling most parents or others who know the child well to complete it independently in 15 to 30 minutes. The CBCL/4-18 has been translated into Dutch. Its good reliability and criterion related validity have been documented thoroughly in the USA (Achenbach, 1991b), and in the Netherlands (Verhulst, Koot, & Akkerhuis, & Veerman, 1990).

Only the problem items part of the CBCL/4-18 was used in the follow-up.

Based on principal components analyses, Achenbach (1991b; see also Achenbach, 1991a) composed eight scales that have common items for the CBCL, the TRF, and the YSR: *Withdrawn*, *Somatic Complaints*, *Anxious/Depressed*, *Social Problems*, *Thought Problems*, *Attention Problems*, *Delinquent Behavior*, and *Aggressive Behavior*. To indicate that the items common to the instruments denote hypothetical variables, the scales are called *cross-informant syndrome constructs*. For some cross-informant constructs, some items that are in one instrument but not in others (e.g. *Disobedient at home*) but that loaded on the construct only for this instrument, are included in the cross-informant syndrome construct only for that instrument. Furthermore, for the CBCL a *Sex problems* scale was composed that was only found for this instrument. For each instrument, the scales have the same items for all age groups covered. The scales have been normed for each instrument and age group separately. Achenbach (1991a,b) also identified *Internalizing* and *Externalizing* groupings of behavioral/emotional problems by performing second-order factor analyses of the eight

syndrome scales. The Internalizing grouping consists of the problem items of the Withdrawn, Somatic Complaints, and Anxious/Depressed scales. The Externalizing grouping consists of the problem items of the Delinquent and Aggressive Behavior scales. As reported by Achenbach (1991a), one-week test-retest reliabilities for these scores range from .82 to .95 (Mean r=.89), and internal consistency coefficients (Cronbach's alpha) for age groups 4-11 years range from .56 to .96 (Mean by z transformation=.83) for boys, and from .54 to .96 (Mean by z transformation=.84) for girls.

Parent Interview II (PI-2)

To obtain parental information on family background variables and child problems at home, we conducted a second Parent Interview (PI-2). The interview included items concerning family constellation, health, behavioral outcome, the need of professional help, and the use of services. The interview also included questions concerning perinatal factors which were used to construct indices of correlates of problem ratings at Time 1. The information from PI-2 concerning family background variables was used to construct indices of problem correlates for Time 2, following the procedure described in Chapter 8. Information on child problems obtained from PI-2 was also used to construct a measure of home problems (see Chapter 8 for details).

Problems at school

Teacher's Report Form (TRF)

To obtain teacher's reports of the children's behavioral/emotional problems, we used the problem item part of the Teacher's Report Form (TRF; Achenbach, 1991c). This part has the same format as the problem items part of the CBCL/4-18. Ninety-three items of the TRF have exact counterparts in the CBCL, the other items have been written to cover problem behavior that may only be observable in a school environment (e.g., *67. Disrupts class discipline*). The competence part of the TRF mainly covers behavior related to formal learning, which is not applicable to 4-5-year-olds (the computer scoring program doesn't even compute a TRF competence score for ages 4-5). This part of the checklist was omitted from our study, and replaced by a more age-appropriate instrument to tap preschool-related competence, the NOSP.

The TRF has been translated into Dutch. Its good reliability and criterion related validity have been documented thoroughly in the USA (Achenbach, 1991c). Reliability and criterion related validity for the Dutch version have been reported by Verhulst and Akkerhuis (1986). As reported by Achenbach (1991c), two-week test-retest reliabilities for these scores range from .82 to .96 (Mean r=.92), and internal consistency coefficients (Cronbach's alpha) for age groups 5-11 years range from .70 to .97 (Mean by z transformation=.89) for boys, and from .63 to .97 (Mean by z transformation=.89) for girls.

Nijmegen Observation Scales for Preschoolers (NOSP)

The Nijmegen Observation Scales for Preschoolers (NOSP; [Nijmeegse Observatieschalen

voor Kleuters]; Rost, 1992) were designed to tap preschoolers' level of school-related competence. The instrument consists of two sections. The first section includes items to rate cognitive abilities required for following formal education. The second section consists of 7-point Likert-type items concerning social-emotional competence. We only used this second section as an alternative to the TRF competence section to be used as a correlate of parent and teacher ratings of problem behavior and significant problems. The items from this section are summarized into four scales: *Task-related behavior* (11 items); *Social behavior* (17 items); *Affect* (6 items); and *Self-help* (9 items). Rost (1992) obtained internal consistencies (Cronbach's alpha) for these scales of .90, .92, .88, and .73, respectively. Rost (1992) found that all scales loaded on one factor, explaining 71.7% of the total variance in the instrument.

Teacher's problems ratings

Teachers filled out a questionnaire containing questions regarding aspects of the behavior of the child that constituted a problem or reason for worry, need for professional help, and the level of impairment caused by the behavioral/emotional problem. Answers to these questions were used as outcome variables in the follow-up study reported in Chapter 8.

Language problems

Language Screening Instrument (LSI)

To assess language problems at preschool age we used the Language Screening Instrument (LSI, [Taal Screening Instrument]; Gerritsen, 1988). The LSI consists of three parts: a formal test of the child's language competence (37 items for 4-year-olds, and 39 items 5-year-olds), which is supplemented by a parent and a teacher rating scale (12 and 6 items, respectively). The test part measures active and passive vocabulary, verbal comprehension, and syntax. The parent and teacher questionnaire ask for information on language use, comprehension and development. Test-retest reliabilities of the instrument were .75-.78. Internal consistency of the parent questionnaire was .74, and of the teacher questionnaire .84. Sensitivity and specificity of the instrument with formal language testing as criterion were 98.5% and 89%, respectively. The parent and teacher rating scales were used in the follow-up study.

Temperament

Dimensions of Temperament Survey-Revised (DOTS-R)

We translated the Dimensions of Temperament Survey-Revised (DOTS-R; Windle & Lerner, 1986), which is a 54-item, factor-analytically derived instrument that measures nine temperament attributes, and is suitable for children from the preschool level (self-report versions exist for children from the late school age level and for young adults). The DOTS-R is designed to be completed by a parent or another caregiver who has extensive experience with the child. A 4-point, Likert-scale response format is used for each item ranging from *usually false* (1) to *usually true* (4). Temperament attribute scores are formed by summing

DOTS-R item scores in accordance with the person's 1 through 4 responses. Fourteen items are reversed before summing items to form subscale scores. High subscale scores indicate higher levels of each attribute. The nine temperament attributes assessed are: *Activity Level-General* (7 items), *Activity Level-Sleep* (4 items), *Approach-Withdrawal* (7 items), *Flexibility-Rigidity* (5 items), *Mood* (7 items), *Rhythmicity-Sleep* (6 items), *Rhythmicity-Eating* (5 items), *Rhythmicity-Daily Habits* (5 items), and *Task-Orientation* (8 items). Windle and Lerner (1986) reported that in a sample of 114 preschool children (mean age 4.12 years), internal coefficients for the above nine DOTS-R attributes are .84, .87, .84, .79, .91, .80, .80, .70, and .79, respectively. The factorial structure of the DOTS-R for preschoolers has been confirmed in a sample of 234 Japanese preschool children (Windle, Iwawaki, & Lerner, Submitted). Significant relationships have been reported for adolescents between the DOTS-R Young Adult version attributes and measures of perceived social and cognitive competence (Windle et al., 1986), stressful life events and mental health (Windle, 1989), and measures of verbal and performance intelligence (Matheny, 1989).

Parenting stress

Nijmegen Parenting Stress Index (NPSI)

The Nijmegen Parenting Stress Index (NPSI, [Nijmegen Ouderlijke Stress Index]; De Brock, Vermulst, & Gerris, 1990) assesses the level of perceived parental stress originating from several child and parent characteristics within the caregiving context. The NPSI is a modified Dutch version of Abidin's (1983) Parenting Stress Index, which originally consisted of 101 items in 13 subscales. The items are scored by the parent on a 6-point Likert scale ranging from *completely disagree* (scored 1) to *completely agree* (scored 6). We used a short form which includes 25 items (14 new or modified), that are derived from scales measuring the perceived child characteristics: demandingness (5 items), acceptation (3), mood (2), distractability (2), adjustment (1), and reinforcement (1), and the perceived parent characteristics: competence (6), depression (3), attachment (1), and health (1). These items had high loadings on one general parenting stress factor. De Brock, Vermulst, and Gerris (1990) reported a high internal consistency for the short form of the NPSI (Cronbach's alpha is .95 for mothers, and .94 for fathers). The instrument discriminates well between parents of children referred to outpatient mental health services and parents of non-referred children (De Brock, Vermulst, & Leenders, 1990).

Life-events

Life Events Questionnaire (LEQ)

To assess potentially stressful life-events, parents filled out a questionnaire containing the 32 items from the questionnaire used by Berden (1992), and one open-ended item to indicate life-events that were not mentioned in the questionnaire. Only items were included that stated or implied a negative event, and items that were not spuriously related to problem behavior scores. The items had a yes/no response format to indicate whether or not an event had

occurred during the 12 months before the Time 2 assessment. The item scores (0,1) were summed into a total life-events score. Events mentioned in the open item were scored present if they met the inclusion criteria and added to the total life-events score.

CHAPTER 4

Internal validity of the CBCL/2-3

The general aim of the study reported in this and the next chapter was to provide information on the psychometric properties of the CBCL/2-3. The specific aims were:
(1) to study the factor structure of CBCL/2-3 narrow-band and broad-band constructs in different samples;
(2) to compose scales of the items that measure these constructs, and to compute the internal consistency reliability of these scales;
(3) to compute the test-retest reliability and interrater agreement of the scales;
(4) to perform Dutch-American comparisons on CBCL/2-3 scales;
(5) to study the convergent and discriminant validity of the scales; and
(6) to study the criterion related validity of the scales.

4.1 Factorial structure

The validation of instruments for the assessment of psychopathology involves the investigation of two types of hypotheses (Morris, Bergan, & Fulginiti, 1991). The first has to do with the extent to which the items selected to assess a given construct (such as aggression or anxiety) actually do measure that construct. The second has to do with the relationships between the constructs being assessed and other variables. This chapter considers the first type of hypotheses.

To examine the structure of CBCL/2-3 syndrome constructs, factor models were studied. In clinical assessment, constructs are generally measured by scales that quantify variation in the level of the constructs being assessed. The construct is treated as a quantitative dimension. Using an assessment scale to measure, for example, the level of anxiety in a child, the location on the dimension is represented by a test score on this measure. However, the quantitative dimension being assessed is not directly observable. Rather, it is a hypothetical construct (or in statistical language "latent variable") whose presence must be inferred from observable responses to test items. These observable responses may be thought of as manifest indicators of the latent variable. One approach to validating the assumption that a latent variable can be inferred from a set of observed variables is factor analysis.

Achenbach (Achenbach, 1992; Achenbach, Edelbrock, & Howell, 1987) used principal components analysis (PCA) to identify, what he called "syndromes", i.e.,

groups of "problems that tend to occur together, without implying any particular model for the nature or causes of disorders" (Achenbach, 1992, p.10) for the CBCL/2-3. Although the results of PCA are generally similar to the results of principal factor analysis (PFA), when large numbers of items (e.g., greater than 40) are used (Snook & Gorsuch, 1989), we chose to analyze our data using a principal factor model both in initial exploratory analyses and in subsequent confirmatory analyses.

Several statistical considerations exist for preferring the common factor model over the principal components model as an exploratory procedure (Gorsuch, 1983). In the social sciences few variables are thought to be error-free and it is seldom expected that all variances will be predicted from the factors. Therefore, PCA which makes these assumptions may not be the method of choice. Further, PCA generally produces higher loadings than PFA. Loadings may be sufficiently inflated as to be misleading. Finally, PFA automatically becomes a PCA when that is the best model, i.e., in case the communalities are actually unities. In the reverse case, where communalities are less than one, these communalities are seldom as low as would be given by a PFA.

In performing factor analysis on the CBCL/2-3, we are confronted with the general problem of comparing factors from the same variables applied to different populations. This issue may be illustrated by the results obtained by Achenbach when he performed PCA to the CBCL/2-3 items in two different samples in 1986 (Achenbach, Edelbrock, & Howell, 1987) and in 1992 (Achenbach, 1992). In both samples six components were obtained and scales were composed of the items loading above a certain cutpoint, labeled Social Withdrawal or Withdrawn, Depressed or Anxious/Depressed, Sleep Problems, Somatic Problems, Aggressive (Behavior), and Destructive (Behavior). Despite similar labels, the proportion of items contained in each of the 1987 scales that reappeared in the 1992 scales was only .36, .13, .88, .75, .47, and .57, respectively. Conversely, of the items in the 1992 scales the following proportions of items were also in the 1987 versions: .36, .18, 1.0, .62, 1.0, and .73. Thus, the invariance of components subsumed under the internalizing heading was poor across samples. The moderate to high overlap among items in the Sleep Problems, Aggressive Behavior, and Destructive Behavior scales reflect that in the 1992 versions these consist largely of subsamples of items contained in the 1987 versions. Considering that the CBCL/2-3 may be of value as an instrument to assess psychopathology in rather diverse populations, our purpose was to compose CBCL/2-3 scales based on robust factors, i.e., based on factor solutions that are reasonably invariant across variations in the selection of subjects. To the extent that invariance can be found across changes in subjects, the factors have a wider range of applicability as generalized constructs. The factors would then be applicable to several populations and could be expected to generalize to other similar populations as well. We replicated our analyses across three independent samples: children referred to mental health agencies; children from the general population; and pairs of monozygotic and dizygotic twins.

We used exploratory factor analysis (EFA) to obtain sets of items that load consis-

tently on recurring factors. These sets of items were used to specify the factor models to be evaluated in a confirmatory factor analysis (CFA). In the present study Jöreskog's (1971) congeneric measurement model was fitted to the data by using the mainframe version of the computer program LISREL 7 (Jöreskog & Sörbom, 1989). The congeneric measurement model assumes that a subject's score on an item depends on his score on the underlying factor and an error of measurement. Items are allowed to have different factor loadings and error variances. Factors and measurement errors are assumed to be uncorrelated, and the mean measurement error is assumed to be zero.

4.1.1 Outline of data analysis

Samples and instrument

The exploratory and confirmatory factor analyses were performed on the scores on CBCL/2-3 items obtained in the clinical (N=458) and community (N=420) sample. Because we wanted the clinical sample to be representative of children who were referred for behavior problems and developmental delays *of unknown origin*, we excluded from the analyses the 32 children who were diagnosed as mentally retarded, leaving 426 children in the clinical sample.

We had the opportunity to replicate our analyses once more on a twin sample consisting of 1306 three-year-old twin pairs (total N=2612) who participated in a study on the influence of genetic factors on the development of behavior problems in young children (Van Den Oord, 1993).

Subjects in the twin sample were 1306 pairs of 3-year-old twins (1291 boys, 1321 girls) from a target sample of 1892 pairs (73% response rate). The twins' mean age was 42.1 months (SD = 4.0). Employment rate was 98% for fathers, and 29% for mothers. The mean occupational level (Van Westerlaak, Kropman, & Collaris, 1975) of mothers was 3.60 (SD = 1.37), and of fathers 3.51 (SD = 1.40). The mean maternal age was 33.0 years (SD = 3.9), and the mean paternal age was 35.6 years (SD = 4.6).

In the Netherlands, about 85% of the parents of all newborns are paid a home visit by a commercial organization which promotes certain products. During this home visit parents of twins are asked to participate in the twin register kept by the Department of Psychonomics of the Free University of Amsterdam. Forty percent of all multiple births in the Netherlands are registered. CBCLs for ages 2-3 were mailed to parents of three-year-old twins. Non-responders were sent reminders and contacted by telephone. For 73% of the twin pairs both parents filled out one CBCL/2-3 for each child. For 20% only maternal ratings were available. For 8% only paternal ratings were available. Questions about demographic characteristics were contained in the questionnaire.

Both parents completed the CBCL/2-3 for most children in the twin sample. However, there was no check on the independence of completion by each parent. Therefore, it was not obvious which of both checklists should be included in the analyses. In order to be able to include all subjects the item scores were averaged across parents in

those cases in which both parents had filled out the checklist, after ascertaining that mothers' and fathers' ratings did not differ systematically (see Van Den Oord, 1983).

The factor analyses were done on each sample separately.

Exploratory factor analysis of syndrome constructs

We first performed principal factor analyses with promax rotation using the SAS (1989) statistical package. In this promax rotation the orthogonal solution (in our case varimax) is used as a basis for estimating an ideal, oblique solution. The unrotated factor matrix is rotated to the best least-squares fit to this target solution using weights obtained from the varimax rotation. To create a better solution than that given by the orthogonal rotation, the moderate and low loadings need to be lower than in the orthogonal solution while the high loadings remain relatively high. Rotating to the varimax and promax criteria is a procedure recommended by several authors (e.g., Gorsuch, 1983; Jöreskog, 1978). The procedure gives a good simple structure. Furthermore, it provides a basis for determining whether orthogonal or oblique factors are more acceptable. If the correlations among the oblique factors are negligible, then the orthogonal varimax solution would be accepted as a resonable solution. If the correlations are significant, the oblique solution would be the choice.

After performing principal factor analysis using unweighted least squares, we subjected the first 5 to 12 factors from the analyses to varimax and subsequently to oblique promax rotations. We examined the 5- to 12-factor rotated factor solutions to identify sets of items that consistently grouped together, along the following guidelines. First, factors were sought that replicated across the 5 to 12 rotated factor solutions in each sample. Second, items were sought which loaded highly on only one factor. Third, we avoided factors too narrow in scope, i.e., having one or two items with a very specific content. Fourth, similar factors were sought across samples, consisting of the same set of items manifesting salient loadings on the one factor in each sample.

The items loading $\geq .30$ on the 5 to 12 rotated factors obtained in each sample were listed side-by-side to identify the version of each factor that included the maximum number of high loadings, which also loaded highly on the other versions. We selected the solution with the best factorial structure according to two following criteria: the rotated solution with the highest proportion of items that consistently recurred in the versions was selected; if we could not decide using the first criterion, the rotated solution with the highest loadings and the fewest cross-loading items was retained.

After identifying the best solution in each of the samples, items that were common to the corresponding factors from two or more samples were listed. Sets of items that had loadings on corresponding factors in at least two of the three samples were used to specify the factor model to be evaluated in the CFA.

Confirmatory factor analysis of syndrome constructs

An initial model was based on the sets of items obtained from the EFA in the three samples. A loading from an item on a construct was specified when the item was included

in the set of items. When the item was not included in the set of items the loading was fixed at zero. All correlations between the constructs were estimated. In subsequent steps we examined whether this initial model had to be elaborated with additional parameters, because its pattern of factor loadings was very restrictive and it did not allow correlated errors of measurement. The respecifications of the model in these steps were guided by statistics (modification index [MI] and expected parameter change [EPC]) LISREL can provide. Every time a respecified model was fitted to the data we also examined whether it could be simplified. Factor loadings (standardized) less than .10 and error of measurement correlations less than .10 were considered non-substantial, and therefore fixed at zero the next time the model was fitted. This procedure of estimating new parameters and fixing non-substantial parameters continued until the overall fit indices did not improve anymore.

When ordinal variables are to be analyzed, as is the case with CBCL/2-3 items, analyzing ordinary product-moment correlations results in underestimates of factor loadings and in overestimates of unique variances (Jöreskog & Sörbom, 1988). In this case it would be appropriate to carry out the analysis starting with a matrix of polychoric correlations. Since we wanted to compare the factor structures obtained in three different samples we kept as many items as possible in the analyses. Some of these had quite low frequencies, especially in the community and twin samples. Unfortunately, this prohibited the computation of polychoric correlations because iterations failed to converge. Therefore, unweighted least squares were used as input.

Model evaluation

Several criteria can be used to evaluate a factor model. A first criterion is that the model should be simple. A more parsimonious model is preferable to a model with more parameters. Further, a model should be interpretable and parameter estimates should be acceptable. For instance, the occurrence of improper solutions (e.g., negative error variances, correlations which are larger than one) can be indicative of a misspecified model. Finally, a model must account for the observed correlations/covariances. To assess how well a model approximates the data, several goodness-of-fit indices have been proposed. The most frequently used index of absolute fit is (an approximation) of the chi-square test statistic because it offers a statistical test of the validity of the model. However, this index is not applicable here, because unweighted least squares estimations were used, instead of maximum likelihood or weighted least squares as would be appropriate. Furthermore, the nonnormal distribution of CBCL/2-3 scores, especially in nonclinical samples, prohibited other adjustments that would allow the use of the chi-square test.

Fit indices that were reported were the goodness-of-fit index (GFI), the adjusted goodness-of-fit index (AGFI), and the root of the mean squared residuals (RMR). The GFI and the AGFI are based on a comparison of the observed correlations with the correlations predicted by the factor model. A value of one indicates a perfect fit, while a value close to zero indicates a bad fit. Estimating more parameters results in increases in the fit of a model. AGFI adjusts for this phenomenon. The RMR is the root of the mean

squared residuals which can be interpreted when, as in this study, correlations are used as input. Fit indices for three models were reported. A baseline model, which assumes that no common factors underlie the items and correlations between the items are therefore zero; the initial model; and the model which resulted from respecifying this initial model. Fit indices for the baseline model can be used to get an impression of the lower bounds of the fit indices. In our study, both the initial model and the respecified model were fitted to three different samples. Differences between the fit indices of these models in the samples were used to evaluate the appropriateness of the respecifications.

4.1.2 Results of the factor analyses of syndrome constructs

Exploratory factor analysis of syndrome constructs

Problem items that are scored as present for extremely low or extremely high percentages of children in a sample do not differentiate well and should be removed from a test. Two items were reported for less than 5% in the two community samples as well as in the clinical sample, and were therefore excluded from the analyses: *39. Headaches (without medical cause)*, and *57. Problems with eyes without medical cause*. No items were reported to be present for 95% or more for more than one of the samples, leaving 97 items in the analyses.

In the clinical sample as well as in the community sample, the first seven factors found in the 7- through 9-factor solutions had nearly exactly the same items loading $\geq .30$ on similar factors in consecutive rotated solutions. These factors replicated quite well in seven factors from the 9-factor solution of the twin sample. Inspection of the factor inter-correlation matrix (see Appendix D.1) for the oblique factors showed that the oblique solution was clearly preferable to the orthogonal solution. The correlations among factors were low to moderate in all three samples. The moderate and low loadings obtained from the oblique rotations were lower and fewer loaded $\geq .30$ than in the varimax rotations, while the high loadings were similar in both rotations. Moreover, considerably fewer cross-loadings appeared in the oblique than in the varimax rotations, which improved the interpretability of the factors.

The items included in the seven-factor solutions in the clinical and community sample and those in the comparable factors from the nine-factor solution in the twin sample with loadings $\geq .30$ were listed side-by-side. In this way we could identify items that would constitute the constructs to be evaluated in the CFA. Appendix D.2 contains the loadings (standardized regression coefficients) of the 69 items with loadings $\geq .30$ on the seven corresponding factors obtained from the promax rotations in the three samples. These 69 items were included in the initial factor model to be evaluated in the CFA. Based on the items included in these factors we applied the following preliminary labels to the factors: Oppositional; Withdrawn/Depressed; Aggressive; Anxious; Overactive; Sleep Problems; and Somatic Problems. Fourteen items had a cross-loading of .30 to .40 in one of the samples. Of these, item *56. Clumsy* had cross-loadings above .30 in two samples.

Confirmatory factor analysis of syndrome constructs

An important constraint on the measurement model is that it must not be underidentified. This constraint permits estimates of unknown parameters, such as factor loadings. This constraint is met when the number of unknown parameters in the model is less than the number of degrees of freedom. Based on the number of variances and covariances in our model, we had a total of 2415 degrees of freedom. In the initial model, a total of 83 unknown parameters had to be estimated: 69 factor loadings plus 14 cross-loadings of items for which loadings were not constrained because cross-loadings $\geq .30$ were obtained for these items in the EFA in one of the three samples. Loadings of items which were not expected to influence the construct were constrained to equal zero, and the loading of a single item on each construct to equal 1.0. Because relatively stable estimates may be expected with a 4:1 subjects-to-parameters ratio (Tanaka, 1987), samples sizes were judged to be adequate. We used oblique rotations.

Table 4.1
Model fit indices for the CBCL/2-3 syndrome constructs

Model	df	GFI	AGFI	RMR
Clinical sample				
Baseline	2346	.274	.253	.194
Initial	2242	.901	.894	.072
Final	2207	.932	.926	.059
Community sample				
Baseline	2346	.380	.362	.153
Initial	2242	.917	.910	.056
Final	2207	.937	.931	.049
Twin sample				
Baseline	2346	.269	.247	.197
Initial	2242	.955	.951	.049
Final	2207	.971	.968	.039

Note. An unweighted least squares estimation procedure was used. The number of items was 69. df = degrees of freedom; GFI = Goodness of Fit Index; AGFI = Adjusted Goodness of Fit Index; RMR = Root of Mean Squared residuals. Sample sizes were for the clinical sample N=426, for the community sample N=420, and for the twin sample N=1306.

Table 4.1 presents the fit indices of the seven-factor baseline model, initial model, and final model for the 69 items that were left after the exploratory analyses. The initial model was a large improvement over the baseline model. All fit indices indicated that the final model offered the better description of the test structure. The AGFI indicated that this better fit was not only because more parameters were estimated in this model. The higher values of the fit indices for the final model in all three samples may be regarded as an indication of the validity of the respecifications of the initial model. The fit indices for the initial as well as for the final model were somewhat higher in the twin sample than in the clinical and community samples. However, this need not indicate that the model had a better fit in the first sample. As indicated earlier, this difference could also reflect the considerably smaller size of the latter two samples. Also, the mean of both parents' ratings may have provided more reliable scores.

Table 4.2 displays the factor loadings obtained from fitting the final model to the item correlations in the three samples. Only cross-loadings with an absolute value $\geq .30$ are shown. In general, parameter estimates were quite acceptable. However, during the process of respecification of the model, the loading of item *5. Can't concentrate* on Overactive exceeded the value of 1.00 in all three samples, while it should be between 1.00 and -1.00. This unacceptable high loading was accompanied by a negative loading $> .30$ on Oppositional. To further respecify the model guided by the modification indices the loading of item 5 on Overactive was fixed at .95. The final model included 21 factor correlations, 69 measurement errors, and 17 correlated errors of measurement; 101 parameters had to be estimated: 68 factor loadings (having the loading of item 5 constrained at .95) and 33 cross-loadings.

The interpretation of most factors is clear. Factor I is defined by high loadings of the items *81. Stubborn*, *16. Demands must be met*, *83. Sulks*, *85. Temper tantrums*, *44. Angry moods*, *88. Uncooperative*, *97. Whining*, and *13. Cries much*, reflecting oppositional and demanding behavior, and lack of emotional regulation. This factor may be labeled Oppositional. Factor II was labeled Withdrawn/Depressed, because highest loading items were *71. Little interest*, *98. Withdrawn*, *43. Looks unhappy*, *67. Unresponsive*, *23. Doesn't answer*, and *26. No fun*. Aggressive behavior against people and objects characterizes the items with high loadings on Factor III, such as *40. Hits*, *53. Attacks people*, *35. Fights*, *17. Destroys own things*, and *18. Destroys other's things*, and thus was entitled Aggressive. Two items, *20. Disobedient*, and *91. Too loud*, that loaded on the Oppositional factor in the exploratory analyses migrated to the Aggressive factor in the CFA. The items loading high on Factor IV, *92. Upset by new*, *73. Shy*, *87. Too fearful or anxious*, *10. Clings to adults*, and *3. Afraid to try new things* reflect anxious, fearful, and shy behavior, suggesting the label Anxious. Factor V was clearly defined by high loadings of *5. Can't concentrate*, *6. Can't sit still*, and *59. Quickly shifts activity*, which may be adequately covered by the label Overactive. Factor VI may be entitled Sleep Problems, consisting of items that all have to do with sleep disturbances, with high loadings

Table 4.2
Factor loadings for CBCL/2-3 items obtained from confirmatory factor analyses in the clinical, community, and twin sample

	Factor / Items	Clinical sample		Community sample		Twin sample	
	I. Oppositional (17 items)						
	8. Can't wait	.397		.453		.631	
	13. Cries much	.571		.507		.590	
A	15. Defiant	.567	-.365 [4]	.459		.367	.346 [3]
A	16. Demands must be met	.735		.677		.786	
A	29. Easily frustrated	.551		.434		.499	
A	30. Easily jealous	.551		.433		.557	
AD	33. Feelings easily hurt	.494		.384		.553	
D	36. Gets into everything	.479		.377		.465	
A	44. Angry moods	.680		.557		.671	
A	66. Screams	.493		.382		.414	
A	69. Selfish	.306		.313		.389	
W	81. Stubborn	.817		.715		.886	
A	82. Moody	.545		.379		.550	
	83. Sulks	.641		.665		.793	
A	85. Temper tantrums	.668		.565		.704	
A	88. Uncooperative	.564		.556		.643	
AD	96. Wants attention	.515		.432	.423 [5]	.461	
A	97. Whining	.588		.538		.617	
	II. Withdrawn/Depressed (10 items)						
W	2. Acts too young	.583		.393		.337	
W	23. Doesn't answer	.423		.499		.567	
W	26. No fun	.574		.406		.354	
AD	43. Looks unhappy	.563		.663		.486	
	56. Clumsy	.259		.291		.169	
W	67. Unresponsive	.520		.455		.360	
W	70. Little affection	.390		.541		.285	
W	71. Little interest	.633		.631		.481	
	76. Speech problem	.127		.326		.248	
	77. Stares blankly	.456		.529		.391	
	80. Strange behavior	.384		.448		.415	
W	89. Underactive	.351		.284		.280	
AD	90. Sad	.276	.319 [4]	.352		.495	
W	98. Withdrawn	.600	.367 [1]	.631		.416	.364 [4]
	III. Aggressive (9 items)						
D	14. Cruel to animals	.483		.393		.333	
D	17. Destroys own things	.629		.502		.588	
D	18. Destroys other's things	.631		.455		.580	
A	20. Disobedient	.383	.353 [5]	.311		.397	
A	35. Fights	.666		.570		.653	
A	40. Hits	.717		.607		.661	
D	42. Hurts accidentally	.521		.534		.496	
	53. Attacks people	.728		.580		.611	
A	91. Too loud	.421		.271	.368 [1]	.527	

Table 4.2 (Continued)

Factor / Items	Clinical sample		Community sample		Twin sample	
IV. Anxious (9 items)						
3. Afraid to try new things	.518		.539		.511	
W 4. Avoids eye contact	.058	.418 [2]	.354		.550	
AD 10. Clings to adults	.584		.517		.586	
21. Disturbed by change	.413		.334		.377	
32. Fears	.442		.303		.316	
AD 37. Upset by separation	.525		.423		.613	
AD 68. Self-conscious	.477		.431		.482	
AD 73. Shy	.589		.606		.720	
AD 87. Too fearful or anxious	.590		.526		.590	
92. Upset by new	.668		.631		.730	
V. Overactive (5 items)						
D 5. Can't concentrate	.950		.950	-.303 [1]	.950	-.329 [1]
6. Can't sit still	.834		.787		.790	
11. Constantly seeks help	.494		.540		.606	
D 59. Quickly shifts activity	.759		.668		.849	
W 62. Refuses active games	.461		.472		.368	
VI. Sleep Problems (7 items)						
SL 22. Doesn't want to sleep alone	.612		.452		.390	
SL 38. Can't sleep	.705		.568		.572	
SL 48. Nightmares	.527		.514		.592	
SL 64. Resists going to bed	.710		.575		.554	
SL 74. Sleeps little	.538		.529		.444	
SL 84. Talks or cries in sleep	.451		.416		.467	
SL 94. Wakes often	.726		.701		.622	
VII. Somatic Problems (3 items)						
SO 1. Aches	.661		.578		.608	
SO 12. Constipated	.421		.145		.271	
SO 45. Nausea	.394		.358		.467	
SO 52. Painful bowel movements	.255		.330		.262	
SO 78. Stomachaches	.484		.423		.579	
SO 93. Vomiting	.246		.125		.294	

Note. Loadings are unweighted least squares LISREL estimates. Items that were retained in the scales are printed in bold type. Cross-loadings are given followed by the number of the factor [in brackets] on which the cross-loading occurred. Superscripts indicate items that are comprised in the CBCL/2-3 syndrome scales constructed by Achenbach (1992): A = Aggressive; W = Withdrawn; D = Destructive; AD = Anxious/Depressed; SL = Sleep Problems; SO = Somatic Problems.

on items such as *94. Wakes often*, *38. Can't sleep*, and *64. Resists going to bed*. The items with highest loadings on Factor VII concern physical complaints such as *1. Aches*, and *78. Stomachaches*. This factor was labeled Somatic Problems.

Results were quite similar for the three samples. Items with relatively high factor loadings on a factor in one sample also had relatively high loadings on the same factor in the other samples. Pearson correlations and RMRs were computed for the pairwise comparisons of factor loadings between samples as measures of the congruity of the constructs across samples (cf. Tanaka & Huba, 1974; Tanaka, 1987). The results are reported in Table 4.3 and indicate high mean congruity for all scales except Withdrawn/Depressed and Anxious. The congruity for Withdrawn/Depressed was low for

all comparisons. For Anxious the mean congruity was low for the clinical versus twin sample comparison. Although the mean congruity for Sleep Problems was quite acceptable, the congruity was low for the clinical versus twin sample comparison.

Table 4.3
Pearson correlation coefficients (PCC) and Root of the Mean Squared Residuals (RMR) for pairwise comparisons of estimates of factor loadings between samples

Factor	Clinical/ Community		Clinical/ Twin		Community/ Twin		Mean across samples	
	PCC	RMR	PCC	RMR	PCC	RMR	PCC	RMR
Oppositional	.875	.094	.833	.084	.945	.118	.898	.099
Withdrawn/Depressed	.417	.109	-.143	.158	.237	.134	.180	.134
Aggressive	.901	.118	.818	.077	.688	.110	.820	.102
Anxious	.755	.114	.448	.173	.785	.107	.690	.131
Overactive	.971	.055	.925	.079	.892	.100	.940	.078
Sleep Problems	.787	.095	.467	.126	.947	.063	.810	.095
Somatic Problems	.999	.063	.871	.076	.847	.108	.965	.082

Table 4.4
Intercorrelations of CBCL/2-3 factors obtained from confirmatory factor analysis

Clinical sample	1.	2.	3.	4.	5.	6.	7.
1. Oppositional	1.000						
2. Withdrawn/Depressed	.431	1.000					
3. Aggressive	.681	.329	1.000				
4. Anxious	.401	.638	.158	1.000			
5. Overactive	.703	.390	.633	.242	1.000		
6. Sleep Problems	.442	.121	.217	.320	.294	1.000	
7. Somatic Problems	.278	.350	.170	.354	.169	.307	1.000

Community sample	1.	2.	3.	4.	5.	6.	7.
1. Oppositional	1.000						
2. Withdrawn/Depressed	.519	1.000					
3. Aggressive	.500	.332	1.000				
4. Anxious	.444	.534	.069	1.000			
5. Overactive	.712	.467	.465	.299	1.000		
6. Sleep Problems	.285	.250	.162	.326	.293	1.000	
7. Somatic Problems	.344	.195	.079	.206	.254	.432	1.000

Twin sample	1.	2.	3.	4.	5.	6.	7.
1. Oppositional	1.000						
2. Withdrawn/Depressed	.600	1.000					
3. Aggressive	.715	.487	1.000				
4. Anxious	.540	.638	.265	1.000			
5. Overactive	.754	.536	.684	.425	1.000		
6. Sleep Problems	.489	.356	.372	.347	.395	1.000	
7. Somatic Problems	.351	.387	.232	.304	.277	.384	1.000

Table 4.4 shows the pattern of estimated factor intercorrelations for the three samples. This pattern was fairly similar across the clinical and community samples. The range of factor intercorrelations for the clinical sample was .121 to .703, with a mean of .363. The range of factor intercorrelations for the community sample was .069 to .712, with a mean of .341. For the twin sample the range of factor correlations was .232 to .754, with a mean of .454, being somewhat higher than in the other two samples. This may be a consequence of the larger size of the twin sample, producing less biased estimates than in the smaller clinical and community samples. This pattern also indicates some relatively high intercorrelations in all three samples, e.g., among the oppositional, aggressive, and overactive constructs, suggesting the existence of higher order factors underlying the obtained factors.

Syndrome scales were composed to be used as valid representations of the constructs. Inclusion of an item in a scale was guided by the following criteria. First, an item was only included in the scale of a construct when its loading on that construct exceeded .30 in two of the three samples. Such an item was considered a valid indicator of the construct. Second, the item was not allowed to have loadings on more than one other construct above an absolute value of .30. In this case the item may also be considered a valid indicator of another construct. Just as constructs should be different from each other, this is desirable for their scales too. Finally, the item had to improve the reliability of the scale. This was determined by computing Cronbach's alpha for the scale with and without the item. Using these criteria, the following items had to be deleted from a scale. Item *15. Defiant* was deleted from the Oppositional syndrome. Items *56. Clumsy*, *76. Speech problem*, *89. Underactive*, and *98. Withdrawn* were deleted from the Withdrawn/Depressed scale. Item *32. Fears* was deleted from the Anxious syndrome. Finally, three items had to be deleted from the Somatic Problems scale: *12. Constipated*, *52. Painful bowel movements*, and *93. Vomiting*. In two cases the decision to decide for or against deletion was not straightforward. We considered to retain item *98. Withdrawn* in the Withdrawn/Depressed scale because it may be regarded as one of the defining items for the factor. However, since both cross-loadings were positive, and since the scale was sufficiently defined by the other items, we decided against it. Item *5. Can't concentrate* was retained in the Overactive scale despite two cross-loadings above cutpoint, because it was clearly the prime defining variable, and both cross-loadings were negative.

Table 4.5 reports the reliabilities of the syndrome scales. The reliabilities were acceptable for all scales except for Somatic Problems. The low alpha for this scale was not surprising, because it consisted of only three items. Reliabilities for the Withdrawn/Depressed scale were acceptable for the clinical and community samples, but only moderate for the twin sample.

Table 4.5
Internal consistency coefficients (Cronbach's alpha) of the CBCL/2-3 scales

Scales	Clinical sample	Community sample	Twin sample
Oppositional	.901	.861	.910
Withdrawn/Depressed	.735	.732	.638
Aggressive	.848	.756	.819
Anxious	.792	.757	.828
Overactive	.773	.770	.775
Sleep Problems	.807	.738	.701
Somatic Problems	.504	.430	.593

It might be argued that the composition of the syndrome scales could have been affected by the inclusion of data from both twins in the twin sample, or by the use of the mean paternal ratings. Indeed, data obtained for pairs of twins are not independent. As a consequence, standard errors and tests of significance are incorrect, and should not be used in analyses when data from both twins are included. Estimates of the factor loadings, however, are correct (see Goldstein, 1987; Weng, 1990). To assess whether the estimates of the factor loadings were affected by the interdependence of the twin data, we fitted the factor model separately to maternal ratings of the first twin, maternal ratings of the second twin, paternal ratings of the first twin, paternal ratings of the second twin, maternal ratings of both twins, and paternal ratings of both twins. Of the 101 loadings estimated from each of the analyses of maternal and paternal ratings in the sample including both twins, 97 and 98 loadings, respectively, were within the range of loadings obtained from each of the analyses of maternal and paternal ratings in the samples including only one member of the twin pair.

By computing the mean of the parental ratings, it was assumed that mothers and fathers observe the child's behavior in similar situations, and that they share a common understanding of the behavioral descriptions. Loadings obtained from the separate analyses of the maternal and paternal ratings of both twins were highly correlated with the loadings obtained from the analyses using the mean of maternal and paternal ratings (r = .996 and .992, respectively). If only paternal ratings had been used, one additional item (*15. Disobedient*) would have been included in the Oppositional scale, because the cross-correlation of this item on the Aggressive scale was below .30 using paternal ratings. Use of only maternal ratings would have had no effect on the composition of the scales. Other analyses of the twin data also yielded negligable differences in means and variances of the maternal and paternal ratings (Van Den Oord, Koot, Boomsma, Verhulst, & Orlebeke, submitted), and provided support for the assumption stated above (Van Den Oord, 1993).

4.1.3 Second-order groupings of syndromes

In multivariate analyses of problem behavior in school-aged children and adolescents, two broad-band dimensions of behavior have been found consistently (see Achenbach & Edelbrock, 1978). Achenbach (1966) named them internalizing versus externalizing behaviors, which have now become generally accepted terms (see Cicchetti & Toth, 1991). Similar dimensions have been found in multivariate analyses of preschoolers' problem behaviors. These have been labeled apathy/withdrawal versus anger/defiance (Kohn & Rosman, 1972), aggressive/hyperactive versus anxious/withdrawn (Fowler & Park, 1979; Hoge et al., 1985; Tremblay et al., 1987), and conduct/restless versus emotional/miserable (Richman, Stevenson & Graham, 1982). By applying second-order factor analysis to the six syndrome scales found for the CBCL/2-3, Achenbach (Achenbach, 1992; Achenbach, Edelbrock, & Howell, 1987) obtained two principal factors, designated as Internalizing and Externalizing. The syndrome scales Anxious/Depressed and Withdrawn were grouped under Internalizing, the Aggressive Behavior and Destructive Behavior scales under Externalizing. Sleep Problems and Somatic Problems were not strongly enough associated with either second-order factor to be included in either of the groupings. We tried to obtain similar second-order factors for the seven syndrome scales that resulted from the factor analyses reported above by applying exploratory factor analyses to the scales in all three samples, followed by confirmatory analyses.

The pattern of estimated factor intercorrelations was fairly similar across the three samples (see Table 4.4). In all three samples there were some relatively high intercorrelations among the Oppositional, Aggressive, and Overactive syndromes, and among the two syndromes Withdrawn/Depressed and Anxious. This pattern suggested the presence of two broad-band dimensions.

To study the higher-order factor structure for the seven syndrome scales, we first applied exploratory factor analyses with varimax/promax rotations to the observed intercorrelations (see Appendix D.3) for each of the three samples using unweighted least squares. One- two-, and three-factor higher-order solutions were specified. In all three samples, the three-factor solution was interrupted by the program because no solution could be obtained. Both the one-factor and the two-factor solutions were replicated very well across samples (see Appendices D.4 and D.5).

To choose which model to prefer for representing the higher order factor structure, one- and two-factor models were specified in LISREL. Because there were only seven observed variables, weighted least squares estimation could be used to perform an accurate χ^2 test. To perform an accurate χ^2 test in the twin sample, we randomly selected one twin from each twin pair.

The one-factor model had 14 degrees of freedom. This model yielded the following fit indices: clinical sample, $\chi^2 = 133.81$ ($p = .000$), AGFI = .847; community sample, $\chi^2 = 49.81$ ($p = .000$), AGFI = .882; twin sample, $\chi^2 = 166.01$ ($p = .000$), AGFI = .820. The χ^2 tests indicated that the one-factor model had to be rejected for all three

samples.

For the two-factor model, the varimax/promax rotation was implemented in LISREL. From the varimax/promax rotation one reference variable was chosen for each higher order factor (Jöreskog, 1978). The variable with the highest loading on one factor and the lowest loading on the other factor was chosen as a reference variable. In the LISREL model the loading of this reference variable was estimated for the factor on which it loaded high, and fixed at zero for the other factor. In this case, Aggressive and Anxious were chosen as reference variables for the first and second factor, respectively.

The two-factor model, which had 8 degrees of freedom, yielded the following fit indices: clinical sample, $\chi^2 = 35.11$ ($p = .000$), AGFI = .930; community sample, $\chi^2 = 19.83$ ($p = .011$), AGFI = .918; twin sample, $\chi^2 = 21.37$ ($p = .006$), AGFI = .959. Fit indices indicated that the two-factor model fitted better to the data than the one-factor model. However, the χ^2 test sugggested that the fit of the two-factor model also was not entirely satisfactory. In addition, the two-factor model yielded a negative error variance for the Anxious syndrome in the twin sample. To avoid biased estimates of the other parameter the error variance of Anxious was fixed at .00 (cf. Gerbing & Anderson, 1987).

The first factor was defined by high loadings in every sample for Aggressive (mean of the standardized loadings $M = .747$), Oppositional ($M = .736$), and Overactive ($M = .644$), and may be labeled as an Externalizing grouping. The second factor was defined by high loadings for Anxious ($M = .883$) and moderate loadings for Withdrawn/Depressed ($M = .461$). This factor may be labeled as an Internalizing grouping. The largest cross-loading was for Withdrawn/Depressed on the Externalizing grouping ($M = .272$). The mean loading of Sleep Problems or Somatic Problems never exceeded .294 on either the Externalizing or Internalizing dimension.

The Pearson correlations between Internalizing and Externalizing raw scores were .39, .40, and .53 in the clinical, community, and twin sample, respectively. This reflects the fact that children who have elevated scores in one of the two areas also tend to have somewhat elevated scores in the other area.

4.2 Test-retest reliability

To assess the reliability of the CBCL/2-3 syndrome scales, the CBCL/2-3 was completed twice by 51 respondents (49 mothers, 2 fathers) randomly selected from the community sample. We computed Pearson correlations between the ratings for 26 boys (11 2-year-olds; 15 3-year-olds) and 25 girls (14 2-year-olds; 11 3-year-olds) over a mean interval period of 19.4 days (SD=6.6). As shown in Table 4.6, test-retest rs of the CBCL/2-3 scores ranged from r=.94 for total problems in girls to r=.59 for the Withdrawn/Depressed scale in girls (mean r=.82). All rs were significant at $p < .001$ and above .75, except for the Withdrawn/Depressed scale in both boys (r=.62) and girls (r=.59), and for the Somatic Problems scale in boys (r=.59). Of the 10 comparisons between the reliability coefficients of boys and girls only one yielded a significant diffe-

rence. Girls' total problems score seems to be more reliable than boys' (*rs* .94 vs. .75, $z=2.56$, $p=.01$). However, this one difference could be expected by chance in the 10 comparisons that we made, using a $p<.01$ protection level (Sakoda, Cohen, & Beall, 1954).

Because it is not only important to know rank-order stability of the scale scores as expressed by the Pearson correlation, but also to know changes in magnitude of the scores over a short interval, we computed *t* tests between the same ratings for boys and girls separately. Ten scores declined significantly ($p<.05$). Three of the significant declines would be expected by chance in the 20 comparisons that we made, using a $p<.05$ protection level. The differences that yielded the smallest *t* values are indicated in Table 4.8, because these were the most likely to be significant by chance.

The tendency for reported problems to decline over test-retest intervals has also been found in other studies of preschool rating scales and questionnaires (e.g., Achenbach, 1992; Richman & Graham, 1971). However, we found a clear sex difference in declines. Over the 3-week interval, for boys all scores except Aggressive and Somatic Problems showed significant declines ($p<.05$). For girls, however, only the Anxious scale score and the total problem score declined significantly.

Table 4.6
Three-week test-retest reliability of CBCL/2-3 scale scores

Scale	Boys (n=26)	Girls (n=25)	Mean *r* (n=51)
Oppositional	.91[a]	.86	.88
Withdrawn/Depressed	.62[ab]	.59	.60
Aggressive	.80	.89	.85
Anxious	.81[a]	.84[ab]	.83
Overactive	.85[a]	.83	.84
Sleep Problems	.76[ab]	.76	.76
Somatic Problems	.59	.83	.73
Internalizing	.79[a]	.84	.81
Externalizing	.89[a]	.92	.90
Total Problems	.75[a]	.94[ac]	.87
Mean *r*	.79	.85	.82

Note. All *rs* significant at $p<.001$. [a] Time 1 > Time 2, $p<.05$ by *t*-test.
[b] Time 1 vs. Time 2 difference not significant when corrected for number of comparisons.
[c] Significant sex difference at $p<.01$.

4.3 Interparent agreement

To assess interparent agreement, we computed Pearson correlations and *t*-tests between scores on CBCL/2-3s independently completed by both parents for 48 cases randomly selected from the community sample, supplemented with CBCL/2-3s independently completed by both parents for 12 consecutive cases at the outpatient child psychiatry unit of the Sophia Children's Hospital. Most correlations between parents' scores were significant at $p < .01$ (see Table 4.7), except for Somatic Problems in boys ($r = .12$), and Withdrawn/Depressed ($r = .23$), Anxious ($r = .26$), and Internalizing ($r = .29$) in girls. For Sleep Problems, the interparent agreement was significantly higher for girls than for boys ($z = 2.48$, $p < .05$). However, in 10 comparisons two may be expected by chance, using a .05 protection level (Sakoda et al., 1954). No significant differences in levels of scale scores between mothers and fathers were found.

Table 4.7
Interparent agreement on CBCL/2-3 scale scores

Scale	Boys ($n=30$)	Girls ($n=30$)	Mean *r* ($n=60$)
Oppositional	.71	.57	.64
Withdrawn/Depressed	.49	.23[a]	.37
Aggressive	.52	.60	.56
Anxious	.59	.26[a]	.44
Overactive	.62	.65	.64
Sleep Problems	.48	.84[b]	.70
Somatic Problems	.12[a]	.48	.31
Internalizing	.64	.29[a]	.48
Externalizing	.67	.65	.66
Total Problems	.63	.69	.66
Mean *r*	.56	.56	.56

Note. All correlations were significant at $p < .01$, except those indicated by [a].
[b] Significant sex difference at $p < .05$.

4.4 Dutch-American comparison

As described above, the only other studies that factor-analyzed CBCL/2-3 data were those reported by Achenbach, Edelbrock, and Howell (1987), and Achenbach (1992). Despite the not more than partial overlap of items comprised in the syndromes replicated in these studies, correlations between syndrome scores for 1987 and 1992 versions of the scales were very high for the five scales Sleep Problems (.99), Aggressive Behavior (.97), Destructive Behavior (.93), Internalizing (.93), and Externalizing (.99), high for Somatic

Problems (.85), and substantial for Anxious/Depressed (.73) and Withdrawn (.75) (Achenbach, 1992). The very high and high correlations may be explained by the fact that the 1992 scales were to a large extent shortened versions of the 1987 scales, in which apparently the "core" items, i.e., those that may be regarded as a particularly robust representation of the core construct (cf. Achenbach, Verhulst, Baron, & Althaus, 1987) were maintained. This was considerably less the case with the Anxious/Depressed and Withdrawn scales. The 1992 Anxious/Depressed scale had only two items in common with its 1987 counterpart. Five items were the same in both versions of the Withdrawn scale. This change in items implies a somewhat different meaning for both constructs which is reflected in Achenbach's (1992) adaptation of the scale labels from Depressed into Anxious/Depressed and from Social Withdrawal into Withdrawn. Because comparability of methods for the assessment of psychopathology across studies is important for integrating findings from different studies into a cumulative knowledge base, and for detecting cross-cultural differences and similarities in psychopathology (cf. Achenbach et al., 1987), we compared the syndromes obtained in the present study with those reported by Achenbach (1992).

Table 4.8
Number of items in common and Pearson correlations between Dutch and American versions of CBCL/2-3 scales

Dutch scales	American scales	*N* of Items in Common	Correlation *N*=420
Oppositional	Aggressive Behavior	9	.94
Withdrawn/Depressed	Withdrawn	6	.88
Aggressive	Destructive Behavior	4	.82
	Aggressive Behavior	4	.80
Anxious	Anxious/Depressed	5	.84
Overactive	–	–	–
Sleep Problems	Sleep Problems	7	1.00
Somatic Problems	Somatic Problems	3	.56
Internalizing	Internalizing	14	.90
Externalizing	Externalizing	20	.97

Note. All correlations were significant at $p < .001$.

As a quantitative test of the similarity of the syndromes across cultures, for each child in the matched referred and nonreferred sample, scores were computed by summing the raw scores of 0, 1, and 2 on each CBCL/2-3 item belonging to each American syndrome scale. Pearson correlations were computed between the children's raw scores on the Dutch and American syndrome scales. Table 4.8 gives the correlation coefficients between scores on the Dutch and American scales that were the most similar, as well as

the number of corresponding items in these scales. The correlations in this table were all significantly higher than those between any other combinations of Dutch and American syndromes (except Internalizing and Externalizing), with the smallest $z=2.50$ ($p<.01$) for the difference between the highest r and the second highest r for each syndrome.

The *Oppositional* scale showed a very high correlation of .94 with the American Aggressive scale with nine items common to both versions. One common item (*15. Defiant*) was part of the initial factor model, but was deleted from the Oppositional scale because it had two cross-loadings above cutpoint. Two items in the Oppositional scale were included in the American Anxious/Depressed scale, while two other items were included in the American Withdrawn scale, and one in the Delinquent scale.

The *Withdrawn/Depressed* scale had a very high correlation with the American Withdrawn scale ($r=.88$). Six items were common to both versions of the scale. Unfortunately, two other common items (*89. Underactive* and *98. Withdrawn*) could not be retained in the Dutch scale, the first because of low loadings obtained in the CFA, the second because of cross-loadings above the cutpoint.

The *Aggressive* scale had high correlations with the American Destructive scale as well as with the American Aggressive scale of .82 and .80, respectively. It had four items in common with the American Destructive scale, of which three were among the four highest loading items of this latter scale. It also had four items in common with the American Aggressive scale, of which, again, three were among the four highest loading items of this latter scale.

The cross-national correlation for the *Anxious* and Anxious/Depressed scales was .84. The scales had five items in common. Item *4. Avoids eye contact* of the Anxious syndrome was included in the American Withdrawn syndrome, although it should be noted that for the Dutch clinical sample it also had a loading above cutpoint (.42) on Withdrawn/Depressed, but not on Anxious.

The Dutch 5-item *Overactive* scale did not have an American counterpart at all. Two of the three most defining items of this scale (*5. Can't concentrate* and *59. Quickly shifts activity*) were among the items with low loadings on the American Destructive scale. Further, one item (*62. Refuses active games*) was part of the American Withdrawn scale. The highest correlation of the Overactive scale was with the American Destructive scale ($r=.75$).

The *Sleep Problems* scale was perfectly replicated across countries ($r=1.00$) with all seven items belonging to both versions of the scale.

The correlation between both versions of the *Somatic Problems* scale was low ($r=.56$), compared to the cross-national correlations reported above. All six items in the initial Dutch Somatic Problems factor were part of the American Somatic Problems scale. Only three out of the six in the initial model were retained in the final scale, however. Therefore, a low correlation might be expected.

The broad-band *Internalizing* and *Externalizing* scales showed considerable cross-national similarities with respective correlations of .90 and .97. The two versions of the Internalizing scale had 14 items in common, while 20 items were common to the two

versions of the Externalizing scale. Of the Dutch Externalizing scale, the following five were included in the American Internalizing scale: *33. Feelings easily hurt*; *62. Refuses active games*; *81. Stubborn*; *88. Uncooperative*; *96. Wants attention.*

4.5 Discussion

Dimensions of problem behavior

The findings reported in this chapter show that the CBCL/2-3 is a parent rating scale that enables a reliable distinction between different types of problem behavior in children 2-3 years of age. Seven constructs were identified in exploratory and confirmatory factor analyses: Oppositional, Withdrawn/Depressed, Aggressive, Anxious, Overactive, Sleep Problems, and Somatic Problems. High congruity was found for the factor solutions across three different samples, except for Withdrawn/Depressed. The interpretation of these constructs may be as follows.

The oppositional construct reflects negative emotionality, lack of emotional control, lack of cooperation, and demandingness on the part of the child. Some of the items in the Oppositional scale that refer to negative emotionality and intensity (e.g., *13. Cries much*, *82. Moody*, and *85. Temper tantrums*) are reminiscent of Bates' (1980, 1987) description of difficult temperament. However, 'Difficultness' is not an appropriate label for the items in the oppositional scale, because, first, this scale comprises a much broader scope of problem behaviors and second, agreement of the conceptual definition of difficult temperament is lacking. In the view of other temperament researchers (e.g., Thomas & Chess, 1977), apart from high intensity and negative emotionality, a difficult temperament may also include arrythmicity, withdrawal, and low adaptability. A label that is broad enough and yet distinguishing the behavior from hostile aggression and conduct disorder is "oppositional". The DSM-III-R (APA, 1987) and ICD-10 (WHO, 1989) diagnostic guidelines for Oppositional Defiant Disorder include many of the behaviors covered by the Oppositional construct. These include: loss of temper; defiance of adult requests or rules; deliberate annoyance of other people; angry and resentful behavior; easily being annoyed by other people; low frustration tolerance. They exclude behavior violating laws and the basic rights of others, such as shown by cruelty, bullying, assault, destructiveness, and theft, which are part of the diagnostic criteria of Conduct Disorder.

The aggressive construct is represented by items that all except two (*20. Disobedient* and *91. Too loud*) reflect aggressive acts to individuals (including animals) and properties. Aggression may be defined as any act of behavior that involves the deliberate infliction of injury, harm, or discomfort (e.g., Olweus, 1979; Parke & Slaby, 1987). Aggressive behavior is an acknowledged feature of Conduct Disorder as defined by the ICD-10 and DSM-III-R classification systems.

The overactive construct items reflect concentration problems, short attention span, and overactivity. A hyperactivity-inattentiveness factor has consistently been found for children from age 6 onwards using parent as well as teacher ratings (see Taylor, 1988). Until now, a separate attention problems or hyperactivity construct had not been identified

for 2-3-old children.

The behaviors that define the anxious construct mainly reflect stranger anxiety and separation anxiety. Both types of anxiety may be considered as part of normal development for toddlers and preschool children. Beyond this age these may indicate the presence of emotional disorders such as Separation Anxiety Disorder (DSM-III-R and ICD-10), the DSM-III-R Avoidant disorder, or the ICD-10 Social Sensitivity Disorder.

The withdrawn/depressed construct is defined by items reflecting social unresponsiveness, and immaturity, as well as depressed mood. A similar pattern of behaviors emerged for some children in studies of interactions among children in peer groups (e.g., see Rubin, Hymel, Mills, & Rose-Krasnor, 1991), and it has been identified as a factor in multivariate analyses of problem behavior ratings (e.g., Barkley, 1988). However, it has no clear counterpart in any diagnostic classification system.

Sleep problems, defined by various behaviors including sleep disturbance as well as resisting sleep constitute a well-known problem for many of the parents of young children as well as for clinicians and consultants (e.g., Douglas & Richman, 1984).

Somatic problems in the present analyses reflect general pains, stomaches, and nausea. Somatic problems as a factor obtained from multivariate studies generally covers a much broader range of physical complaints (e.g., Achenbach, 1991a).

The two second-order constructs, comprising the anxious and withdrawn/depressed constructs, and the aggressive, oppositional, and overactive constructs, indicate that the broad-band groupings of Internalizing and Externalizing problem behavior may be already identified at a very early age. Clusters of problems such as feeding problems and toileting problems, that are frequently mentioned in the clinical literature for this age group (e.g., Richman & Lansdown, 1988), were not found in the present analyses, despite the presence of five items in the checklist regarding each of these problem areas.

It must be stressed that interpretations of factors are post hoc unless a hypothesis testing procedure is used (Gorsuch, 1983). In this study confirmatory factor analysis was used to evaluate a factor model that was specified using results of preliminary exploratory factor analyses. However, since the model was tested on the same data from which it was derived, this essentially leaves us in a post hoc situation. Therefore, the interpretation of the factors that were obtained may not be regarded as final but only as a lead for further research.

Reliability of the scales

The internal consistency and test-retest stability coefficients of the CBCL/2-3 scale scores were quite acceptable for most scales. Internal consistency was low for Somatic Problems, which may be expected with a scale composed of very few items. The test-retest reliability of the Withdrawn/Depressed scale was only moderate. Interparent agreement for the scale scores was generally moderate, but quite comparable with the level of agreement found for the CBCL/2-3 by Achenbach (1992), as well as for other cross-informant comparisons (Achenbach, McConaughy, & Howell, 1987). The interparent agreement for scales reflecting behavior in the internalizing spectrum was lower than for

externalizing scales, which agrees with Achenbach's (1992) findings for the 3-year age-group.

Cross-national comparison of CBCL/2-3 factors

Achenbach identified six factors on CBCL/2-3 scores of 2-3-year-olds in two partially overlapping American samples (Achenbach, 1992; Achenbach, Edelbrock, & Howell, 1987). The items included in the American CBCL/2-3 syndrome scales showed high to very high correlations with the scores of the same children on the Dutch CBCL/2-3 syndrome scales. However, despite the considerable statistical congruency between American and Dutch scale scores as expressed by these correlations, most scales have somewhat different interpretations across nationalities.

The Sleep Problems scale was perfectly replicated across cultures. Needless to say, sleep problems constitute a salient problem in young children across cultures.

This is less so with somatic problems. Only three items in the cross-national counterparts of this syndrome were the same. The Dutch version of the Somatic Problems scale comprised only three items, two of which were also high loading items in the American version. However, of six other items included in the American scale, five loaded $\geq .30$ in at least one of our samples, and one loaded .284. One reason for the comparatively low cross-national concordance of this scale may be that most of the items reflecting somatic distress showed a low prevalence among Dutch children. Moreover, many of the Dutch respondents were hesitant to score these items, because they felt uncertain about the medical nature of the complaints. Several of the items on somatic problems in the checklist ask for complaints *without medical cause*.

On the externalizing dimension, the Oppositional scale is highly concordant with the American Aggressive scale, but three out of four items with high loadings on the American Aggressive scale are not in the Oppositional scale. The Dutch Aggressive scale contains equal numbers of items with high loadings on the American Aggressive as well as Destructive scales. The items in the Dutch Aggressive scale almost all imply a violation of the rights of others, most notably physical aggression against persons or properties. The Dutch Overactive scale has no clear American counterpart. The factors that emerged may be ordered along a dimension reflecting increasing levels of disruptiveness. The presence of a second-order Externalizing factor is beyond doubt, given the very high statistical as well as content similarity of the Externalizing scale across cultures.

On the internalizing dimension, withdrawn as well as anxious behaviors are included in scales with reasonable comparibility across cultures. Items reflecting depressed mood, such as *43. Looks unhappy without good reason* and *90. Unhappy, sad, or depressed* are not consistently included in one of both scales across cultures or across samples. These items are variously included in the American 1987 Depressed scale and 1992 Anxious/Depressed scale, and in the Dutch Withdrawn/Depressed scale. The Internalizing factor also seems fairly replicable across cultures. In the American as well as Dutch analyses the Withdrawn and Withdrawn/Depressed scale, respectively, had the lower loadings on the second-order Internalizing factor. In the American analyses the mean loadings of the

Withdrawn scale were even of equal magnitude (.50) on both the Internalizing and Externalizing factor. As Achenbach (1992) remarks, this may be partially accounted for by the inclusion in this scale of several items (such as *81. Stubborn*) that are included in the Externalizing grouping found for older children (e.g., on the CBCL/4-18). Similarly, in the present study, some items (*81. Stubborn* and *88. Uncooperative*) that are included in the American Withdrawn scale are defining of the Dutch Oppositional scale. We thus may conclude that Withdrawn and Anxious behaviors form salient dimensions of behavior as measured by the CBCL/2-3, but that the position of depressed mood on these dimensions is less firmly established. We also may conclude that the second-order Internalizing factor has reasonable cross-cultural validity, although withdrawn behavior is of less value than anxious behavior for defining this factor. The cross-cultural similarity of the majority of the CBCL/2-3 syndrome scales emerged despite differences in samples as well as factor-analytic techniques between the American and Dutch studies.

Despite the obvious similarities two clear differences between the American and Dutch studies should be noted. Unlike Achenbach (1992), we found a distinct Overactive construct. For this age group a separate factor including items indicative of hyperactivity and attention problems has not been found in any other study using rating scales. Although Behar and Stringfield (1974) obtained a distinct Hyperactive-Distractable factor using the PBQ, only one-sixth of their sample was 3 years old. The present analyses thus seem to indicate the possibility of a further differentiation of toddler's externalizing problem behaviors using the CBCL/2-3.

Another difference regards the correlation between the Internalizing and Externalizing scales in this study, which was only .40 in the community sample. Achenbach (1992) reported a correlation coefficient of .75 between these broad-band scales in the American normative sample. This means that the percentage of common variance among these scale scores was 56% for the American versions ot the scales as compared to 16% for the Dutch versions. The correlation between the Dutch CBCL/2-3 broad-band scales is also considerably lower than those reported for the CBCL/4-18 (Achenbach, 1991b).

Comparison with other rating scales

Multivariate analyses have been done on only one other instrument for obtaining parent ratings of toddlers' problem behavior. Richman, Stevenson and Graham (1982) performed a cluster analysis on the Behaviour Screening Questionnaire (BSQ) scores of 183 3-year-olds. They obtained five clusters: Normal; Normal - Nightwet; Disturbed - Widespread sphincter control; Disturbed - Nightwet - Conduct; and Disturbed - Restless - Food fads. None of the constructs derived from the present study seems comparable to these clusters. This may not be surprising given the small number of only 12 items in the BSQ, and the different technique of cluster analysis, which orders individuals instead of variables.

Several of the factors identified in the analyses of parent ratings have also been found in multivariate analyses of broad-band teacher ratings of problem behavior in preschool children. Using the Preschool Behavior Questionnaire with preschoolers aged 3-6 years,

Behar and Stringfield (1974) found a Hostile-Aggressive factor resembling our Aggressive factor, with high loadings on items reflecting behavior inconsiderate of others, fighting, destructive behavior, bullying, and physical aggression. A second factor, Anxious/Fearful, mainly reflected behaviors subsumed under both our Anxious and Withdrawn/Depressed scales. Items with high loadings reflected anxiety, depressed mood, and apathy. Hyperactive-Distractable emerged as a third factor which was highly similar to our Overactive factor, covering inattentive behavior, concentration problems, and physical restlessness. The first two of these factors have been replicated in several studies (Fowler & Park, 1979; Hoge et al., 1985; Tremblay et al., 1987). In these studies, the items in Behar and Stringfield's Hyperactive-Distractable factor had moderate to low loadings on a Hostile-Aggressive factor. In their study of the Symptom Checklist, Kohn and Rosman (1972) found an Apathy-Withdrawal factor of which the high loading items were very similar to those in the Withdrawn/Depressed factor obtained from the present analyses. Their second factor, labeled Anger-Defiance, was characterized by behaviors covered by our Oppositional and Aggressive factors. In a study by McGuire and Richman (1986) on the PBCL, a factor designated as Conduct/Restless/Aggressive, comprised behaviors covered by all three of our scales under the Externalizing heading. Their second factor, Isolated/Immature, was similar to our Withdrawn/Depressed scale. The third, Emotional/Miserable factor, has no clear counterpart to the syndromes obtained in the present study. Thus, in most of these studies, the problem behaviors covered by the three CBCL/2-3 scales Aggressive, Oppositional, and Overactive were subsumed under one single heading. The summation of the three scales into the broad-band Externalizing scale is in line with this. The behaviors covered by the Withdrawn/Depressed and Anxious scales have been subsumed under various headings. In the present study, the items on these scales have been aggregated into the broad-band Internalizing scale. Thus, the 99-item CBCL/2-3 allows a differentiation of preschoolers' problem behaviors that may not be achieved with existing teacher-ratings.

We may conclude that the factor structure of the CBCL/2-3 is fairly invariant across samples as well as across cultures. The high congruency of the factor structure across different samples, and the similarity of most scales across cultures support the differentiation of young children's problem behaviors into constructs reflecting problems that have to do with opposition, aggression, and possibly overactivity, constituting an externalizing dimension, and anxiety and withdrawal, constituting an internalizing dimension. Sleep problems represent a separate construct at this age.

CHAPTER 5

External validity of the CBCL/2-3

In this chapter findings on two aspects of the external validity of the CBCL/2-3 will be reported: construct-related validity and criterion-related validity. Validity refers to the accuracy with which a procedure or a test measures what it is supposed to measure. The validity of a measure may be ascertained in several ways.

Content validity reflects whether a measure's content includes what it is intended to measure or, in this case, to what extent the items in the CBCL/2-3 may be regarded as a reliable sample of problem behaviors of preschool-aged children. This type of validity will not be addressed in our study. The summary of procedures for assembling CBCL/2-3 items given in the manual for the CBCL/2-3 (Achenbach, 1992, pp.2-3) provides a partial answer to the issue of its content validity.

Construct validity refers to theoretically predicted relations between a concept or construct and a network of related concepts (see Cronbach & Meehl, 1955). One approach to the construct validity of a procedure used to measure a certain concept is by testing associations with assessment procedures that measure similar concepts or constructs (*convergent validity*), and with assessment procedures that measure dissimilar constructs (*discriminant validity*). Ideally, the associations should be tested using both the same methods of data collection (e.g., rating scales) as well as different methods (e.g., observations or formal testing) for similar and dissimilar constructs. Put in a multi-trait multi-method matrix the obtained associations would yield evidence for the measurement of a construct across methods (Campbell & Fiske, 1959). Given the number of subjects in this study, however, it was not feasible to use other methods than rating scales as operationalisations of similar and dissimilar constructs.

In this chapter the relations of the CBCL/2-3 scores with the Behavior Checklist (Richman, 1977), as a measure of convergent validity, and with the Minnesota Child Development Inventory (Ireton & Thwing, 1974), as a measure of discriminant validity, will be reported.

Criterion-related validity concerns the degree to which a measure agrees with an independent criterion of what is being measured. As discussed in Chapter 2, there is no diagnostic instrument, system or taxonomy that can be regarded as a proper criterion for validating the CBCL/2-3 syndrome scales for behavioral/emotional problems. In this chapter referral for mental health services was chosen as a measure of "caseness". We analyzed the degree to which each syndrome scale, the total problem score, and the

individual items of the CBCL/2-3 discriminated between criterion groups consisting of children referred for mental health services and nonreferred children matched for sex, age, and socio-economic status of the family.

5.1 Construct validity of CBCL/2-3 scale scores

An indication of the construct validity of the CBCL/2-3 would be if the ratings of problem behavior were more strongly related to a measure which taps the same aspects of functioning, i.e., problem behavior, than with a measure which taps other aspects of functioning, such as general developmental level. The construct validity of the CBCL/2-3 was assessed by determining the concurrent Pearson correlations of the CBCL/2-3 scores with the BCL score and with the MCDI General Development score. Like the CBCL/2-3, both the BCL and the MCDI are parent rating instruments, i.e., all three measures used the same method of data collection. Therefore, relations in the expected direction of CBCL/2-3 scores with BCL and MCDI scores, respectively may be regarded as evidence of construct validity for the CBCL/2-3 (Campbell & Fiske, 1959). We expected that the correlations of CBCL/2-3 scores with BCL scores, measuring level of problem behavior, would be higher than the correlations with MCDI scores, measuring developmental level. Since the BCL was developed for children aged 3 years, the relation between CBCL/2-3 and BCL was only analyzed for children in this age group (n=207).

Table 5.1
Pearson correlations between CBCL/2-3 scale scores and BCL[a] and MCDI

	BCL-score[b]		MCDI General Development	
Scale	Boys (n = 110)	Girls (n = 97)	Boys (n = 198)	Girls (n = 193)
Oppositional	.57	.61	- .07	- .04
Withdrawn/Depressed	.33	.20[de]	- .19[c]	- .10
Aggressive	.38	.41	.01	- .04
Anxious	.40	.28[ce]	- .11[de]	- .15[de]
Overactive	.53	.55	- .16[d]	- .11
Sleep Problems	.36	.40	.18[d]	.09
Somatic Problems	.22[c]	.14[e]	.10	.01
Internalizing	.43	.29[c]	- .17[c]	- .16[d]
Externalizing	.62	.64	- .09	- .07
Total Problems	.69	.61	- .11[de]	- .09
Mean r	.41		-.06	

Note. [a] Only for age 3. [b] All $p < .001$, except those with superscript [c-d]. [c] $p < .01$. [d] $p < .05$. [e] Not significant when corrected for number of analyses.

All correlations between the BCL score and the CBCL/2-3 scale scores and total problem score (see Table 5.1) were significant at $p < .01$ (except for Withdrawn/Depressed, $p < .05$, and Somatic problems, n.s., in girls), with a mean r of .41, ranging from $r = .69$ for Total Problems in boys to $r = .14$ for Somatic Problems in girls, which was the only nonsignificant correlation. The mean rs were not different for boys and girls.

The BCL was originally designed as a screening instrument to identify preschoolers at risk for behavior problems. Of the three-year-olds studied by Richman (1977), 78.5% could be correctly classified using a BCL score of 10 or higher and a clinical rating of mild, moderate, or severe behavior problems as criterion. The CBCL/2-3 might be applied for the same purpose using a clinical cutoff point, e.g., a score between the 82nd and the 90th percentile as proposed by Achenbach (1992). We compared the agreement between scores >85th percentile on the CBCL/2-3 total problem score and BCL scores ≥10. The sensitivity of the CBCL/2-3 with the BCL as criterion was 71.4, while the specificity was 89.3. Of the children with a BCL score ≥10, 28.6% had a CBCL/2-3 score below the 85th percentile point, and 10.7% of the children with a low BCL score had a high CBCL/2-3 score, producing a 12.4% overall misclassification rate. Thus, using the CBCL/2-3 as a screener, the proportion of false negatives was almost three times as high as the proportion of false positives. Having a high BCL score could be predicted from the CBCL/2-3 score for more than 87% of the cases. The relative risk odds ratio (Fleiss, 1981) for having a high BCL score relative to having a high CBCL/2-3 score was 21.0, which means that children who had a high CBCL/2-3 score had 21 times greater odds of having a high BCL score than children who had a low CBCL/2-3 score.

Most of the correlations of CBCL/2-3 scores with the MCDI General Development score were low and nonsignificant (mean $r = -.06$; see Table 5.1). Of the 20 correlations for boys and girls only two were significant at $p < .01$. Boys' Withdrawn/Depressed and Internalizing scores correlated -.19 and -.17, respectively, with the MCDI General Development score. Of the 20 correlation coefficients for boys and girls in each pair of comparisons three significant correlations might be expected by chance, using a .05 protection level (Sakoda, Cohen, & Beall, 1954). These correlations are indicated in Table 5.1.

5.2 Criterion-related validity

In Chapter 2, we discussed the lack of well-established standards for the presence of psychopathology or the need for help in young children. Although not an infallible criterion, clinical referral, i.e., referral to a mental health or child guidance setting may be regarded as a more valid criterion for assessing the criterion-related validity of problem behavior rating scales than other criteria, most notably when large samples are involved, as in our study (see Achenbach & Edelbrock, 1981). In choosing an appropriate statistical procedure for analyzing differences between the CBCL/2-3 scale scores and item scores obtained by referred versus nonreferred children, the nature of referral status as an independent variable should be considered. In an earlier report (Koot & Verhulst,

1991), we performed analyses of covariance on a limited sample of 175 matched pairs of referred and nonreferred children to obtain estimates of the effect of referral status on the level of item scores. Two considerations made us decide to use regression analysis of scale scores and item scores in the present case. First, referral status should not be regarded as an experimental condition, but rather as an external criterion onto which the scale scores and item scores are regressed. Our second consideration is that variances of item scores obtained in a clinical sample may be expected to be very different from these obtained in a community sample. Homogeneity of variances in the samples analyzed is a basic assumption in analysis of variance. Therefore, we chose regression analysis instead of ANOVA to analyze the discriminative power of both the problem scale scores and the item scores.

Following the procedure used by Achenbach (1992), we compared demographically matched referred and nonreferred samples to prevent confounding of possible demographic differences in scores with referral status. To form demographically matched samples, we drew 326 subjects from the clinical sample of 458 subjects who could be matched to 326 subjects from the community sample of 410 subjects (excluding the 10 referred children in this sample). The subjects were exactly matched for sex (198 boys, 128 girls in each group), age (144 of age 2, 182 age 3 in each group), and SES (111 from lower SES-groups, 121 from middle SES-groups, and 94 from higher SES groups). For the nonreferred sample, 99.1% of CBCL/2-3s were completed by mothers, 0.6% by fathers, and 0.3% by other informants, compared to 62.6%, 6.1%, and 30.4%, respectively for the referred sample. Multiple regressions were performed for all problem scales and for the total problems score for each sex separately. As a measure of the magnitude of differences in scores accounted for by referral status, age, and SES, we computed the percent of total variance in each scale accounted for by each of the independent variables after partialling out variables that accounted for more of the variance.

5.2.1 Referral status differences between scale scores

Table 5.2 displays the percentage of variance in each scale accounted for by significant associations with referral status, age, and SES. The percent of variance accounted for by each variable is represented by the semipartial R^2 (or part correlation) obtained from the regression analyses after partialling out the effects of any other independent variables that accounted for more variance in the scale scores. Appendix E.1 shows the actual means and standard deviations of scale scores for the matched referred and nonreferred samples. In Appendix E.2 the mean scores on each scale and the mean total problem score are grouped according to referral status, sex, and age. Although the points in the figures are connected, the outlined pattern should not be interpreted as the longitudinal course of the development of problems in the individual child.

Table 5.2
Percent of variance accounted for by significant ($p < .05$) effects of referral status, age, SES, and sex in scale scores for matched referred and nonreferred samples

		Boys (n=396)			Girls (n=256)		
Scale	Sex	Referral[a]	Age[b]	SES[c]	Referral[a]	Age[b]	SES[c]
Oppositional	–	20	–	3	17	–	3
Withdrawn/Depressed	–	24	2	–	22	–	–
Aggressive	4	15	–	2	13	–	5
Anxious	–	6	2 *	1 *	7	–	–
Overactive	–	18	–	2	14	–	3
Sleep Problems	<1 *	6	1 *	3	12	–	–
Somatic Problems	–	–	2	2	3	2 *	–
Internalizing	–	16	2	<1 *	17	–	1 *
Externalizing	<1 *	23	–	3	19	–	4
Total Problems	–	26	–	3	26	–	3

Note. [a] All problem scores were higher for referred than nonreferred in significant cases at p = .00001, except Somatic Problems in girls where p=.003. [b] All significant age effects reflected higher scores for 3-year-olds, except Sleep Problems in boys which were scored higher for 2-year-olds. [c] All significant SES effects reflected higher problem scores for lower than upper SES. * Effects most likely to be chance findings.

Referral status differences. All problem scale scores were scored higher for referred than for nonreferred children ($p < .00001$), except Somatic Problems in boys. The effect of referral status on Somatic problems among girls was significant at $p < .003$. No effects of such small probability would be expected by chance in 10 analyses. The effects of referral status were higher for boys than for girls on the Oppositional, Withdrawn/Depressed, Aggressive, Overactive, and Externalizing scales, and higher for girls than for boys on the Anxious, Sleep Problems, Somatic Problems, and Internalizing scales. The effects of referral status on Total Problems was equal for both sexes. According to Cohen's (1988) criteria, all effects of referral status could be considered medium (accounting for 13% to 26% of variance), except on the Anxious scale for both sexes, Sleep Problems for boys, and Somatic Problems for girls, which showed small effects of referral status (2% to 13% of variance).

Age differences. Among boys, 3-year-olds had higher scores on scales reflecting withdrawn/depressed and anxious behavior, and somatic problems. Sleep Problems were scored higher for 2-year-old than for 3-year-old boys. Among girls, 3-year-olds scored higher than 2-year-olds only on the Somatic Problems scale. These significant age differences, however, could be considered small. Furthermore, three out of these six, i.e., the differences obtained for Anxious and Sleep Problems among boys, and for Somatic Problems among girls, could be expected by chance alone, using a .05 protection

level (Sakoda et al., 1954).

SES differences. We also found small, but significant SES-differences for most scales, all indicating higher scores for children from lower SES. Withdrawn/Depressed scores were not significantly affected by SES for either boys or girls, while effects of SES on Anxious were very small for boys, and nonsignificant for girls. Also, for girls SES had no significant effect on Sleep Problems and Somatic Problems scores.

For practical purposes, such as clinical decision making, some users may wish to use the CBCL/2-3 scores in a categorical fashion, e.g., by distinguishing children scoring above a certain cutoff point from those scoring below it. As mentioned above, Achenbach (1992) indicated cutpoints for distinguishing categorically between the normal and the clinical or borderline range. This borderline range was introduced in the 1992 version of the CBCL/2-3, because it often yields stronger associations between clinical status and classification of children according to their scale scores. Phi correlations between clinical status and scores trichotomized into a normal, borderline, and clinical range were somewhat higher than those between clinical status and scores dichotomized into normal and clinical range (including borderline) (Achenbach, 1992). The bottom of the borderline range is defined by approximately the 85th percentile on the total problem, Internalizing, and Externalizing score (normalized T-score of 60), and by approximately the 95th percentile on the other scale scores (normalized T-score of 67). The bottom of the clinical range is defined by the 90th percentile on the total problem, Internalizing, and Externalizing score (normalized T-score of 63), and by approximately the 98th percentile on the other scale scores (normalized T-score of 70). To analyze the association between the categorical classifications of having a CBCL/2-3 score in the borderline/clinical range and referral status, relative risk odds ratios (ORs; Fleiss, 1981) were computed. For each CBCL/2-3 score, children from the matched referred and nonreferred samples were classified according to whether they scored in the normal range or in the borderline/clinical range. Then, the odds were computed that children who had a score in the borderline/range on a particular scale were from the referred sample, relative to the odds for children who were not in the borderline/clinical range. The statistical significance of the ORs was evaluated by computing confidence intervals, the lower border of which must exceed 1.0, and by computing chi-squares for the 2 x 2 tables.

Table 5.3 shows the ORs for relations between scale scores in the borderline/clinical range and referral status, and the percentage of referred and nonreferred children who scored in the borderline/clinical range according to the cutpoints indicated above. All ORs were significant according to confidence intervals greater than 1.0. Chi-squares showed that the observed distribution of CBCL/2-3 scores in the borderline/clinical range among referred and nonreferred children is significantly different ($p < .01$) from random distribution. The largest ORs were for Oppositional (OR=12.6) and Withdrawn/Depressed (OR=11.2), indicating that the odds of being in the referred group were more than 12 or 11 times higher for children, whose Oppositional or Withdrawn scale scores were

in the borderline/clinical range, than for children, whose scores were in the normal range, respectively. ORs for Externalizing, Total Problems, one or more syndrome in the borderline/clinical range, and Internalizing and/or Externalizing in the borderline/clinical range were large relative to the ORs for the rest of the scales, being near or above 8.

Table 5.3
Odds ratios and percent of referred and nonreferred samples scoring in the clinical range

Scale	Odds ratio	Percent above borderline	
		Referred	Nonreferred
Oppositional	12.6	38	5
Withdrawn/Depressed	11.2	38	5
Aggressive	6.6	28	5
Anxious	5.2	17	4
Overactive	5.1	30	8
Sleep Problems	5.9	23	5
Somatic Problems	2.3	12	6
Internalizing	6.3	48	13
Externalizing	7.8	60	16
Total Problems	8.7	63	16
≥1 Syndrome in clinical range	8.4	74	25
Internalizing and/or Externalizing in clinical range	9.2	69	19

Note. $N = 652$ referred and nonreferred 2-3-year-olds matched by sex, age, and SES. In all analyses, the proportion of referred children scoring above borderline was significantly greater than the proportion of nonreferred ($p < .01$ for chi-squares; confidence intervals for odds ratios > 1.0).

5.2.2 Referral status differences between item scores

The same 652 referred and nonreferred 2-3-year-olds exactly matched for sex, age, and SES as in the analysis of criterion-related validity of the syndrome scales were used with the item scores. Multiple regressions were performed for all items with referral status, sex, age, and SES as independent variables. Table 5.4 displays the percentage of variance in each item and the total problem score accounted for by significant ($p < .05$) associations with referral status, sex, age, and SES. The percent of variance accounted for by each variable is represented by the semipartial R^2 (or part correlation) obtained from the regression analyses after partialling out the effects of any other independent variables that

accounted for more variance in the scale scores. Using a .05 protection level, 9 out of 101 effects might be significant at $p<.05$ by chance (Sakoda et al., 1954). The effects most likely to be significant at chance are indicated with the superscript "c".

Referral status differences. As is indicated in Table 5.4, 88 of the 100 problem items and the total problem score were significantly ($p<.05$) related to referral status, indicating higher scores for referred children. Items 1, 3, 12, 24, 32, 36, 39, 41, 68, 73, and 95 were not scored higher for referred than for nonreferred children. On item 79, nonreferred children had higher scores than referred children. According to Cohen's (1988) criteria, regression coefficients accounting for 2 to 13% of the variance in the dependent variable are considered small, and coefficients accounting for 13 to 26% of variance are considered medium. Referral status accounted for medium effects on the following items (percent of variance in parentheses): *5. Can't concentrate* (14%); *29. Easily frustrated* (14%); *58. Punishment doesn't change his/her behavior* (13%); *82. Sudden changes in mood or feelings* (14%); *85. Temper tantrums or hot temper* (16%); *88. Uncooperative* (13%); *96. Wants a lot of attention* (14%). Among the significant effects of referral status on individual item scores, 64 could be considered small (2-13% of variance), while 18 could be considered very small (less than 2% of variance). As was already obvious from the analysis of scale score differences reported in section 5.2.1, differences between referred and nonreferred children accounted for 26% of the variance in the total problems score.

Appendix E.3 shows the percentage of item scores 1 and 2, the mean item scores and the standard error of the mean for children grouped by referral status and sex. The prevalence rate of each problem item reported as present (scored 1 or 2) for the children in the matched samples is graphically displayed in Apppendix E.4. The data points in the graphs represent the percentage points for children grouped according to referral status, sex, and age.

Sex differences. Sex differences were found on 29 items. Seven of these differences reflected higher scores for girls, and 22 reflected higher scores for boys. Eight sex effects accounted for 2% or more of the variance. Seven of these indicated higher scores for boys, one indicated higher scores for girls. Of the 22 items on which boys received significantly higher scores, 12 were from the Externalizing syndrome, while only one was from the Internalizing syndrome. All items on which girls scored higher were from neither of these syndromes. The proportion of all externalizing items on which boys scored higher (12 out of 31) was significantly greater than the proportion of all internalizing items on which they scored higher (1 out of 19) ($\chi^2=5.85$, $p<.01$). Six of the seven effects that indicated higher scores for boys and accounted for 2% or more of variance are included in the Aggressive scale.

Table 5.4
Percent of variance accounted for by significant ($p < .05$) effects of referral status, age, and SES in regression on CBCL/2-3 scores for matched referred and nonreferred samples

	Boys (n=396)			Girls (n=256)		
Item	Referral status[A]	Age	SES[C]	Referral status[A]	Age	SES[C]
1. Aches/pains	--	1[O]	--	2*	--	--
2. Acts too young	10	1[O]	--	12	--	--
3. Afraid to try new things	--	1[O]	--	--	--	--
4. Avoids eye contact	4	--	--	--	1[O]*	--
5. Can't concentrate	14	--	--	16	--	1*
6. Can't sit still	11	--	2	4	--	2
7. Can't stand things out of place	--	--	--	3	--	--
8. Wants everything now	12	--	--	8	--	--
9. Chews nonfood	1	--	--	4	4[Y]*	--
10. Too dependent	3	--	2	4	--	1*
11. Constantly seeks help	7	1[O]*	<1*	6	--	2
12. Constipated	--	--	--	--	4[O]*	--
13. Cries a lot	7	--	1	12	--	--
14. Cruel to animals	5	--	1	--	--	3
15. Defiant	1*	--	--	--	--	4
16. Demands must be met now	12	--	2	7	--	1
17. Destroys own things	6	--	1	7	--	2
18. Destroys others' things	9	--	--	7	--	--
19. Diarrhea	1*	1[Y]	--	5	--	--
20. Disobedient	13	--	--	10	--	3
21. Disturbed by change	8	1[O]*	--	9	--	--
22. Doesn't want to sleep alone	--	--	1	3	--	--
23. Doesn't answer	11	1[O]*	--	6	--	--
24. Does't eat well	--	--	--	--	--	--
25. Doesn't get along with other kids	15	--	--	8	2[O]*	2
26. No fun	4	--	--	5	--	3
27. Lacks guilt	9	--	--	11	--	4
28. Doesn't want to go out	--	--	--	2	--	--
29. Easily frustrated	16	--	--	11	--	1*
30. Easily jealous	6	--	1*	5	--	2
31. Eats or drinks nonfood	4	--	--	2	--	--
32. Fears	--	2[O]	--	--	--	--
33. Feelings are easily hurt	--	--	2	--	--	--
34. Accident-prone	--	--	2	2*	--	2
35. Gets in fights	2	--	3	3	--	3
36. Gets into everything	--	--	--	--	--	--
37. Upset by separation	8	--	1	8	--	2
38. Trouble sleeping	6	1[Y]	--	12	--	--
39. Headaches	--	--	--	--	--	--
40. Hits others	6	--	--	5	--	2

Table 5.4 (Continued)

Item	Boys (n=396)			Girls (n=256)		
	Referral status[A]	Age	SES[C]	Referral status[A]	Age	SES[C]
41. Holds breath	--	1[Y]*	--	--	--	--
42. Hurts accidentally	5	--	--	2	--	--
43. Looks unhappy	8	1[O]	1*	11	--	2
44. Angry moods	5	1[O]	3	7	--	2
45. Nausea	--	--	1	1*	2[O]*	1*
46. Nervous movements	6	--	1	5	--	--
47. Nervous	13	1[O]*	--	9	--	2
48. Nightmares	2	--	1	4	--	--
49. Overeating	1*	--	1	2*	--	--
50. Overtired	4	--	--	5	--	--
51. Overweight	1	--	--	--	--	--
52. Painful bowel movements	--	--	--	2*	--	--
53. Attacks people	7	--	--	6	--	--
54. Picks skin	--	--	1*	--	--	--
55. Plays with sex parts	2	--	2	--	--	2*
56. Clumsy	7	--	--	8	--	--
57. Eye problems	1*	--	--	--	--	--
58. Punishment no behavior change	16	--	--	11	--	--
59. Quickly shifts activity	10	--	<1*	8	--	3
60. Skin problems	1*	--	--	--	--	--
61. Refuses to eat	--	--	--	3	--	--
62. Refuses active games	11	--	1*	7	--	--
63. Rocks head or body	3	--	--	6	--	--
64. Resists going to bed	7	--	2	4	2[O]*	--
65. Resists toilet training	3	2[Y]	--	7	--	--
66. Screams	13	--	2	11	--	2
67. Unresponsive to affection	7	--	1	2*	--	--
68. Self-conscious	--	--	3	--	2[O]*	--
69. Selfish	1*	--	--	--	--	--
70. Little affection	10	--	--	4	--	--
71. Little interest	3	--	--	6	--	--
72. Little fear	2	--	--	7	2[Y]*	--
73. Shy	--	1[O]	--	--	--	--
74. Sleeps little	6	1[Y]*	--	10	--	--
75. Smears bowel movements	5	--	2	5	--	--
76. Speech problem	15	1[O]	--	7	--	--
77. Stares blankly	9	--	<1*	6	--	--
78. Stomachaches	--	--	--	1*	--	--
79. Stores up unneeded things	1[N]*	--	1	--	--	1*
80. Strange behavior	11	--	--	14	--	--

Table 5.4 (Continued)

Item	Boys (n=396)			Girls (n=256)		
	Referral status[A]	Age	SES[C]	Referral status[A]	Age	SES[C]
81. Stubborn	10	--	1	7	--	--
82. Moody	13	2[O]	2	15	--	--
83. Sulks	1*	--	1*	3	--	2*
84. Talks or cries in sleep	--	--	2	1*	--	--
85. Temper tantrums	20	--	2	10	--	--
86. Too concerned with neat or clean	--	--	4	2*	--	--
87. Too fearful or anxious	3	1[O]*	--	5	--	--
88. Uncooperative	16	--	--	9	--	2
89. Underactive	3	--	--	6	--	--
90. Sad	4	1[O]*	--	8	--	--
91. Too loud	10	--	--	7	--	6
92. Upset by new	4	--	--	6	--	--
93. Vomiting	1*	--	--	3	--	--
94. Wakes often	2	2[Y]	1	8	--	--
95. Wanders away from home	--	--	--	--	--	--
96. Wants attention	13	--	--	18	--	1*
97. Whining	11	--	--	8	--	2
98. Withdrawn	9	1[O]*	--	4	--	--
99. Worrying	2	2[O]	1*	--	--	--
100. Other problems	7	--	--	6	--	--
Total Problems	26	--	3	26	--	3

Note. N = 652 referred and nonreferred 2-3-year-olds (396 boys, 256 girls) matched by sex, age, and SES. [A] All significant differences reflected higher scores for referred than nonreferred children, except item *79. Stores up unneeded things*, on which nonreferred had higher scores than referred, indicated by [N]. O = higher scores for older children; Y = higher scores for younger children. [B] All significant differences reflected higher scores for lower SES than for higher SES. * Effects most likely to be chance findings.

Age differences. Age differences were found on 29 items. Nine of these differences reflected higher scores for younger children, and 20 reflected higher scores for older children. Most age effects were very small. The only age difference that accounted for more than 1% of variance was a 2% effect on item *99. Worrying*, on which older children were scored higher. Of the 20 items on which older children received significantly higher scores, three were from the Externalizing syndrome, while seven were from the Internalizing syndrome; 10 were from neither of these syndromes. All items on which younger children scored higher were from neither of these syndromes. The proportion of all internalizing items on which older children scored higher (7 out of 19) was significantly greater than the proportion of all externalizing items on which they scored higher (3 out of 31) (χ^2=3.87, p<.05).

SES differences. SES differences were found on 52 items. Each of these differences reflected higher scores for lower SES children. SES accounted for 2% or more of the

variance in 24 items. Of the 52 items on which lower SES children received significantly higher scores, 27 were from the Externalizing syndrome, nine were from the Internalizing syndrome, while 16 were from neither of these syndromes. The proportion of all externalizing items on which lower SES children scored higher (27 out of 31) was significantly greater than the proportion of all internalizing items on which they scored higher (9 out of 19) (χ^2=7.36, p<.01). SES accounted for 3% of the variance in total problems.

5.3 Discussion

This chapter presented evidence for the construct validity and the criterion related validity of the CBCL/2-3. The convergent construct validity is supported by significant associations between the CBCL/2-3 and the Richman Behavior Checklist using both continuous and categorical scores. We found a mean Pearson correlation of .41 between CBCL/2-3 scores and the BCL score for 207 3-year-olds in the community sample. Higher correlations were found for the Oppositional and Externalizing syndrome. The mean correlation between the BCL and the total problem score was .65. Our findings are consistent with those reported for other samples. Achenbach (1992) reported a Pearson correlation of .58 between the BCL and the CBCL/2-3 total problem scores for 65 children whose parents rated them in a pediatric office, a nursery school, and a study of otitis media. Spiker, Kraemer, Constantine, and Bryant (1992) reported a Spearman correlation of .56 between ratings on both instruments by mothers of a sample of 272 low SES 3-year-old low birthweight children, and a correlation of .77 between nursery school teachers' ratings on the two instruments (N=281). These findings are indicative of convergence between the CBCL/2-3 and the BCL with respect to a general construct of problem behavior among 2- and 3-year-olds (cf. Achenbach, 1992). Furthermore, when used as a screening instrument, results obtained with the two instruments were highly concordant. Classification of the 3-year-olds in the community sample according to the appropriate clinical cutoff points showed a 86.7 percent agreement. Odds ratios showed that children with CBCL/2-3 total problem scores above the cutpoint had a 21 times larger odds of also having a BCL score above the cutpoint than children with total problem scores below the cutpoint. We also found evidence for the discriminant validity of the CBCL/2-3 by computing correlations between the CBCL/2-3 scales and the MCDI General Development score for both ages. The correlations were low with a mean of -.06. Likewise, Achenbach (1992) in a sample of 86 low birthweight and normal children found no significant concurrent correlations between the CBCL/2-3 total problem score and scores on the Bayley Mental Scale at age 2, the McCarthy General Cognitive Index at age 3, and the MCDI total score at both ages.

The evidence of construct validity of the CBCL/2-3 is limited by the absence of other validated measures of behavior problems for young preschoolers that use methods other than parental ratings. As pointed out earlier, to obtain further evidence of construct validity, CBCL/2-3 scores should be compared with scores on other, validated measures

using different methods as well as different informants to measure similar as well as dissimilar constructs. Chapter 7 will provide some more insight into the relation between CBCL/2-3 scores and clinical data obtained from case files.

The criterion-related validity of the CBCL/2-3 was supported by the significant differences between demographically matched referred and nonreferred children's scale scores after effects of the demographic variables age and SES were partialled out. Regression analyses showed medium effects of referral status for both sexes. Small but significant effects of socio-economic status were found, indicating higher scores for boys and girls from lower SES on most scales. Further, small but significant effects of age indicated higher scores for older children on some scales, most notably among boys. Differences between the matched groups were equally obvious using the borderline/clinical cutpoints on the scale scores. The relative odds of being in the referred group were up to 12 times higher for children whose scale scores were in the borderline/clinical range than for children whose scores were in the normal range. In a comparison of 642 demographically matched referred and nonreferred 2-3-year-olds, Achenbach (1992) found a similar pattern of effects of referral status on syndrome scores, but overall these effects were larger than in the present study, most notably for boys. The smaller effects of referral status in the Dutch than in the American study may be partially accounted for by the smaller difference between Dutch referred and nonreferred children's problem scores, as shown in Appendix E.2. For example, the difference between mean total problem scores was 26.6 for the Dutch versus 30.2 for the American matched samples. As in the present study, the American study found age differences almost exclusively for boys. However, differences were in the opposite direction than in the Dutch sample, i.e., higher scores for the younger instead of for the older boys. Further, the present study found SES differences for both sexes, while Achenbach found SES differences only among boys.

The criterion related validity of the CBCL/2-3 was further corroborated by the findings on the total problem score and on the item scores. Referral status accounted for 26% of the variance in the total problem score. Achenbach (1992) found referral status to account for 30 percent of the variance in total problem scores of preschool children aged 2-3, which is quite comparable to our findings. Referral status accounted for 40% of the variance in total problem scores based on parent reports for Dutch children aged 4-16 (Verhulst, Akkerhuis, & Althaus, 1985), for 37% of the variance in scores based on teacher reports (Verhulst & Akkerhuis, 1986), and for 34% in scores based on adolescents' self-reports (Verhulst, Prince, Vervuurt-Poot, & De Jong, 1989). All except one of the significant differences in individual item scores between referred and nonreferred children indicated higher scores for referred children. In the present study the largest effects of referral status (>10 percent) were found for the items *Acts too young*, *Can't concentrate*, *Disobedient*, *Doesn't get along with other kids*, *Easily frustrated*, *Nervous*, *Punishment doesn't change his/her behavior*, *Screams*, *Speech problem*, *Strange behavior*,

Sudden changes in mood or feelings, *Temper tantrums or hot temper*, *Uncooperative*, and *Wants a lot of attention*. Of these 14 items, 11 were among the 17 items for which Achenbach (1992) found the largest effects of referral status, i.e., larger than 11%. The percentages of variance accounted for by the remaining three items (*Acts too young*, *Speech problem*, *Strange behavior*) were in Achenbach's study 5%, 11%, and 11%, respectively. Apparently, across cultures similar behaviors discriminate between children who have and children who have not been referred to mental health services.

The other approach we employed to test the discriminative power of the CBCL/2-3 was to use the 85th percentile of the total problem score in the normative sample to classify children as referred versus nonreferred. The total correct classification rate of 73.3% was lower than the 80.1% found for CBCL total problem scores for a sample of Dutch children aged 4-16 (Verhulst, Akkerhuis, & Althaus, 1985). Although the rate of correct classification seems adequate, it is probably not warranted to consider the total problem score of the CBCL/2-3 as a sole discriminator between referred and nonreferred Dutch children. Thirty-seven percent of the referred children did not have deviant total problem scores, while obviously their parents and caretakers of these children were sufficiently concerned about their behavior or development to seek professional help. Furthermore, of the children having one or more deviant syndrome scores (i.e., above the 95th percentile), only 74.4% would also be classified as deviant using only the total problem score. Not all syndrome scores seem to contribute equally to the total problem score. We computed Pearson correlations between the total problem score and the syndrome scores for all children in the matched samples. The Oppositional syndrome showed the largest correlation with the total problem score (r=.85), while the lowest correlation was found for Sleep Problems (r=.46). Thus, from the correlations between syndrome scores and total problem score, it can be expected that children having deviant Oppositional scores have a much larger chance to also obtain a high total problem score than children having only a deviant Sleep Problems score. Deviant scores for either kind of problems, however, may be sufficient reasons for parents to seek professional help.

CHAPTER 6

Problem behavior in the community sample

The aim of this chapter is to provide prevalence rates of problem behaviors in children aged 2-3 years in the general population, and to identify associated factors of problem behavior at this age. We will provide prevalence rates of behavioral/emotional problems in the sample selected from the general population of the province Zuid-Holland using the CBCL/2-3; identify differences in prevalence rates related to sex, age, and socio-economic status; and provide insight into the association of sociodemographic, child and family factors with problem behavior in this age group.

6.1 Prevalence of individual problem behaviors

The percentages of children in the total community sample who obtained score 1 or 2 on each specific item are given in the first and second columns of Table 6.1.

As presented in Table 6.1, individual behaviors which are thought to be problematic for children aged 2-3 years are reported for quite different numbers of children, ranging from less than 1% for *57. Problems with eyes without medical cause* and *75. Smears or plays with bowel movements* to more than 70% for *8. Can't wait*, *15. Defiant*, *16. Demands must be met immediately*, *20. Disobedient*, *59. Quickly shifts activity*, and *96. Wants attention*. All items were scored for at least one child. All items were scored more often 1 than 2, except *57. Eye problems*, *63. Rocks head or body*, and *100. Other problems*. These items had low frequencies, scored for 6% or less of the children. Although 70 items were scored 1 or 2 for 10% or more of the children, only 18 items were scored 2 in 10% or more of the cases. The item which was most frequently scored 2 was *15. Defiant* (28.1%). Sixteen of the 100 items were scored 1 or 2 for 50% or more of the sample. The behaviors described by these items may be regarded as quite common for children aged 2-3 years. Items scored for less than 10% of the children may be regarded as reflecting relatively uncommon behaviors. The 16 items scored for fifty percent or more of the sample and the 29 items scored for less than 10% of the sample are listed in Table 6.2 according to their association with the internalizing and externalizing syndrome grouping. Of the items that were scored for 50% or more of the children, none was from the internalizing grouping, 13 were from the externalizing grouping, and three were from

Table 6.1
Frequencies of CBCL/2-3 items scored 1 or 2 and percentage of variance accounted for by significant ($p < .05$) effects of sex, age, and socio-economic status in regression on CBCL/2-3 problem scores

Item	% Score 1	% Score 2	Sex[A]	Age	SES[B]
1. Aches	15.2	1.9	–	–	1^{H*}
2. Acts too young	6.2	2.4	1*	–	–
3. Afraid to try new things	27.4	6.0	–	$<1^{O*}$	–
4. Avoids eye contact	22.4	1.4	–	1^{O*}	–
5. Can't concentrate	34.0	11.0	–	–	1^{L}
6. Can't sit still	37.6	22.9	–	–	2^{L}
7. Can't stand things moved	14.0	7.9	2	–	–
8. Can't wait	44.0	26.7	–	–	–
9. Chews nonfood	21.4	5.5	–	1^{Y*}	–
10. Clings to adults	34.3	6.4	–	–	–
11. Constantly seeks help	36.4	4.8	–	–	–
12. Constipated	7.6	2.4	–	$<1^{O*}$	–
13. Cries much	23.1	2.9	–	–	–
14. Cruel to animals	8.3	0.5	–	–	–
15. Defiant	51.7	28.1	–	–	–
16. Demands must be met	54.0	23.6	–	–	1^{L}
17. Destroys own things	26.2	3.6	2	–	–
18. Destroys others' things	19.5	1.4	2	–	–
19. Diarrhea	12.6	3.3	2*	–	–
20. Disobedient	68.6	9.3	2	–	–
21. Disturbed by change	32.4	6.9	–	–	–
22. Doesn't want to sleep alone	12.6	10.2	–	–	–
23. Doesn't answer	38.3	3.3	–	1^{O}	–
24. Does't eat well	35.6	17.6	–	–	1^{L*}
25. Doesn't get along with other kids	12.4	0.7	–	–	–
26. No fun	2.4	0.5	–	–	2^{L}
27. Lacks guilt	33.8	6.9	1*	1^{Y}	–
28. Doesn't want to go out	1.9	1.4	–	–	–
29. Easily frustrated	39.0	7.1	–	–	–
30. Easily jealous	39.8	8.8	–	–	$<1^{L*}$
31. Eats or drinks nonfood	4.5	1.2	–	–	–
32. Fears	36.7	17.6	–	–	–
33. Feelings are easily hurt	41.9	11.7	–	–	1^{L*}
34. Accident-prone	30.5	13.1	–	–	–
35. Fights	22.9	3.3	2	–	1^{L}
36. Gets into everything	42.6	21.2	–	–	–
37. Upset by separation	21.0	4.0	–	–	1^{L*}
38. Can't sleep	16.0	6.0	–	–	–
39. Headaches	3.1	0.2	–	1^{O*}	–
40. Hits others	46.0	4.3	2	–	–

(continued)

Table 6.1 (continued)

Item	% Score 1	% Score 2	Sex	Age	SES
41. Holds breath	4.3	2.4	--	--	--
42. Hurts accidentally	26.4	2.9	--	1^{Y}	--
43. Looks unhappy	5.0	0.2	--	--	--
44. Angry moods	58.6	10.7	1*	--	--
45. Nausea	2.9	0.5	--	2^{O}	--
46. Nervous movements	3.8	1.9	--	--	--
47. Nervous	9.0	0.5	--	--	--
48. Nightmares	18.6	3.3	--	--	--
49. Overeating	5.5	2.1	--	--	--
50. Overtired	17.9	3.1	--	2^{O}	--
51. Overweight	2.4	0.7	--	--	1^{L*}
52. Painful bowel movements	5.0	1.2	--	--	--
53. Attacks people	6.2	0.7	1*	--	1^{L}
54. Picking	23.3	8.6	--	--	--
55. Plays with sex parts	11.2	1.4	--	--	--
56. Clumsy	9.8	2.1	--	--	--
57. Eye problems	0.0	0.2	--	--	--
58. Punishment no behavior change	35.7	12.6	2	--	--
59. Quickly shifts activity	45.5	26.1	--	--	2^{L}
60. Rashes	9.5	4.3	--	--	--
61. Refuses to eat	39.3	7.6	--	--	--
62. Refuses active games	22.4	4.3	--	--	--
63. Rocks head or body	2.9	3.1	--	--	--
64. Resists going to bed	13.6	3.3	--	--	--
65. Resists toilet training	13.1	11.9	1*	5^{Y}	--
66. Screams	35.5	9.5	--	--	--
67. Unresponsive to affection	5.2	1.7	--	--	1^{L}
68. Self-conscious	25.5	1.4	--	1^{O*}	1^{L*}
69. Selfish	32.9	1.7	--	--	--
70. Little affection	7.1	1.0	--	--	--
71. Little interest	4.5	1.7	--	--	--
72. Little fear	25.0	9.0	1*	--	--
73. Shy	42.9	6.7	1^{F*}	--	--
74. Sleeps little	8.3	7.9	--	--	--
75. Smears bowel movements	0.2	0.0	--	--	--
76. Speech problem	9.5	7.4	--	--	2^{L}
77. Stares blankly	9.3	1.4	--	--	1^{L*}
78. Stomachaches	3.8	1.2	--	--	--
79. Stores up unneeded things	14.8	6.9	--	1^{O}	--
80. Strange behavior	2.4	1.2	--	--	--

(continued)

Table 6.1 (continued)

Item	% Score 1	% Score 2	Sex	Age	SES
81. Stubborn	52.4	9.3	--	--	$.1^{L}$
82. Moody	30.2	5.0	--	1^{O*}	--
83. Sulks	45.5	4.5	--	--	--
84. Talks or cries in sleep	24.8	4.0	--	--	--
85. Temper tantrums	41.9	11.7	--	--	--
86. Too concerned with neat or clean	15.5	2.6	--	--	2^{L}
87. Too fearful or anxious	17.4	2.4	--	$<1^{O*}$	--
88. Uncooperative	40.5	2.4	--	--	$<1^{L*}$
89. Underactive	3.1	0.5	--	--	--
90. Sad	6.9	0.0	--	--	1^{L}
91. Too loud	31.7	4.5	--	--	3^{L}
92. Upset by new people or situations	24.5	3.8	--	--	--
93. Vomiting	3.3	0.7	--	--	--
94. Wakes often	26.4	12.4	--	--	--
95. Wanders away from home	6.0	2.1	1^{*}	--	--
96. Wants attention	49.3	21.2	--	--	--
97. Whining	34.5	3.3	--	--	--
98. Withdrawn	5.5	1.4	--	1^{O*}	--
99. Worrying	6.7	0.0	--	1^{O}	--
100. Other problems	1.4	2.1	--	--	--

Note. N = 420 for multiple regression analyses on problem items. ^{A}: All significant sex effects indicated higher scores for males, except *73. Shy*, where females scored higher. ^{O} = Higher scores for older children; ^{Y} = Higher scores for younger children; ^{B}: All significant SES effects indicated higher scores for lower SES, except *1. Aches*, where higher SES scored higher. * = Effects most likely to be chance findings.

neither grouping. Of the items that were scored for less than 10% of the children, eight were from the internalizing grouping, two were from the externalizing grouping, and 19 were from neither grouping. All internalizing low-frequency items were part of the Withdrawn/Depressed syndrome. The proportion of all externalizing items that was scored for more than 50% of the children (13 out of 31) was significantly greater than the proportion of all internaliizing items scored with high frequency (0 out of 19; $\chi^2=7.26$, $p<.01$). Conversely, the proportion of all internalizing items that was scored for less than 10% (8 out of 19) was larger than the proportion of externalizing items scored with low frequency (2 out of 31; $\chi^2=7.34$, $p<.01$)

Table 6.2
Items with high[a] and low[b] prevalence by internalizing and externalizing syndrome group

High prevalence	Low prevalence
Internalizing	
	2. Acts too young
	26. No fun
	43. Looks unhappy
	67. Unresponsive to affection
	70. Little affection
	71. Little interest
	80. Strange behavior
	90. Sad
Neither Internalizing or Externalizing	
15. Defiant	28. Doesn't want to go out
24. Doesn't eat well	31. Eats or drinks nonfood
32. Fears	39. Headaches
	41. Holds breath
	45. Nausea
	46. Nervous movements
	47. Nervous
	49. Overeating
	51. Overweight
	52. Painful bowel movements
	57. Eye problems
	63. Rocks head or body
	75. Smears bowel movements
	78. Stomachaches
	89. Underactive
	93. Vomiting
	95. Wanders away from home
	98. Withdrawn
	99. Worrying
Externalizing	
6. Can't sit still	14. Cruel to animals
8. Can't wait	53. Attacks people
16. Demands must be met	
20. Disobedient	
33. Feelings are easily hurt	
36. Gets into everything	
40. Hits others	
44. Angry moods	
59. Quickly shifts activity	
81. Stubborn	
83. Sulks	
85. Temper tantrums	
96. Wants attention	

[a] Score 1 or 2 for at least 50% of the children.
[b] Score 1 or 2 for less than 10% of the children.

Table 6.3
Items for which significant ($p < .05$) differences between sex, age, and SES effects were found in the community sample by Internalizing and Externalizing syndrome group

Effect	Internalizing	Neither	Externalizing
Sex[A]	2. Acts too young	7. Can't stand things moved	17. Destroys own things
	73. Shy	19. Diarrhea	18. Destroys others' things
		27. Lacks guilt	20. Disobedient
		58. Punishment no behavior change	35. Fights
		65. Resists toilet training	40. Hits others
		72. Little fear	44. Angry moods
		95. Wanders away from home	53. Attacks people
Age[B]	3. Afraid to try new[O]	9. Chews nonfood[Y]	42. Hurts accidentally[Y]
	4. Avoids eye contact[O]	12. Constipated[O]	82. Moody[O]
	23. Doesn't answer[O]	27. Lacks guilt[Y]	
	68. Self-conscious[O]	39. Headaches[O]	
	87. Too fearful or anxious[O]	45. Nausea[O]	
		50. Overtired[O]	
		65. Resists toilet training[Y]	
		79. Stores up unneeded things[O]	
		98. Withdrawn[O]	
		99. Worrying[O]	
SES[C]	26. No fun	1. Aches	5. Can't concentrate
	37. Upset by separation	24. Doesn't eat well	6. Can't sit still
	67. Unresponsive to affection	51. Overweight	16. Demands must be met
	68. Self-conscious	76. Speech problem	30. Easily jealous
	77. Stares blankly	86. Too concerned with neat or clean	33. Feelings are easily hurt
	90. Sad		35. Fights
			53. Attacks people
			59. Quickly shifts activity
			81. Stubborn
			88. Uncooperative
			91. Too loud

Note. [A]: All significant sex effects indicated higher scores for males, except *73. Shy*, where females scored higher. [O] = Higher scores for older children; [Y] = Higher scores for younger children; [B]: All significant SES effects indicated higher scores for lower SES, except *1. Aches*, where higher SES scored higher.

6.2 Sex-, age-, and SES-effects on individual problem scores

To assess the effects of sex, age, and SES on the scores for each problem item in the community sample, multiple regressions were performed for all items with sex, age, and SES as independent variables. Table 6.1 displays the percentage of variance in each item score accounted for by significant ($p < .05$) associations with sex, age, and SES. The percent of variance accounted for by each variable is represented by the semipartial R^2 (or part correlation) obtained from the regression analyses after partialling out the effects of any other independent variables that accounted for more variance in the scale scores. All effects could be considered small, according to Cohen's (1988) criteria. Using a .05

protection level, 9 out of 100 effects might be significant at $p < .05$ by chance (Sakoda et al., 1954). The effects most likely to be significant at chance are indicated with an asterisk. The items for which significant effects were found are listed in Table 6.3 according to their association with the internalizing and externalizing syndrome grouping.

Sex differences. Sex differences were found on 16 items. One of these differences (item *73. Shy*) reflected higher scores for girls, and 15 reflected higher scores for boys. Eight sex effects accounted for 2% of the variance, all indicating higher scores for boys. Of the 15 items on which boys received significantly higher scores, seven were externalizing, and one was internalizing. The proportion of all externalizing items on which boys scored higher (7 out of 31) was not significantly greater than the proportion of all internalizing items on which they scored higher (1 out of 19) ($\chi^2=1.49$, n.s.). Six of the seven effects that indicated higher scores for boys in the Externalizing grouping are included in the Aggressive scale.

Age differences. Age differences were found on 17 items. Four of these differences reflected higher scores for younger children, and 13 reflected higher scores for older children. The only age difference that accounted for more than 2% of variance was a 5% effect on item *65. Resists toilet training*, on which younger children were scored higher. Of the 13 items on which older children received significantly higher scores, two were externalizing, five were internalizing, and 10 were neither internalizing or externalizing. The proportion of all internalizing items on which older children scored higher (5 out of 19) was significantly greater than the proportion of all externalizing items on which they scored higher (1 out of 31) ($\chi^2=3.96$, $p < .05$).

SES differences. SES differences were found on 22 items. Each of these differences reflected higher scores for lower SES children, except *1. Aches*, where from higher SES groups scored somewhat higher. SES accounted for 2% or more of the variance in 6 items. Of the 22 items on which children form lower SES levels received significantly higher scores, 11 were from the Externalizing syndrome, six were from the Internalizing syndrome, and five were from neither of these syndromes. The proportion of all externalizing items on which children from lower SES groups scored higher (11 out of 31) was not significantly greater than the proportion of all internalizing items on which they scored higher (6 out of 19) ($\chi^2=0.03$, n.s.).

6.3 Scale scores in the community sample

Tables 6.4 and 6.5 display the scale means, and the percentage of variance in each scale accounted for by significant associations with sex, age, and SES. The percent of variance accounted for by each variable is represented by the semipartial R^2 (or part correlation) obtained from regression analyses of scale scores on sex, age, and SES. The analyses were performed for the whole sample to detect sex differences, and subsequently for each sex separately to analyse associations with age and SES.

Boys and girls had very similar scale scores and total problem scores. The only sex difference found was for the Aggressive scale, on which boys scored significantly higher than girls. Among boys, 3-year olds had higher scores on the Anxious and Internalizing scales. Among girls, 3-year olds scored higher than 2-year olds on the Somatic Problems scale. Significant effects of SES on scale scores reflected higher scores on the Oppositional, Overactive, Externalizing, and total problem scales in boys from lower SES levels, and higher scores on all scales except Oppositional, Sleep Problems, and Somatic Problems in girls from lower SES levels. Thus, in boys SES was related to the externalizing dimension, whereas in girls SES was related to both the externalizing and internalizing dimension. All significant differences could be considered small. Furthermore, all significant sex and age effects could be expected by chance alone, using a .05 protection level (Sakoda et al., 1954).

Scale scores may also be used in a categorical way to describe the prevalence of behavioral / emotional problems. The prevalence may be presented by the percentage of children having one or more deviant CBCL/2-3 syndrome scores, i.e., scores above the borderline or clinical cutoff point. In Table 6.6 these percentages are listed together with the percentage of children who obtained one or more deviant syndrome scores and/or a deviant total problem score. As in the analyses reported in Chapter 5, the bottom of the borderline range was defined by approximately the 85th percentile on the total problem, Internalizing, and Externalizing score (normalized T-score of 60), and by approximately the 95th percentile on the other scale scores (normalized T-score of 67). The bottom of the clinical range was defined by the 90th percentile on the total problem, Internalizing, and Externalizing score (normalized T-score of 63), and by approximately the 98th percentile on the other scale scores (normalized T-score of 70). Because the syndrome scores are based on fewer items than the total problem score, which may imply somewhat less reliability, a more stringent cutoff point was chosen for the syndrome scores.

In this community sample, 12.6% of the children had one or more deviant syndrome scores using the clinical cutoff criterion, and 21.4% had one or more deviant syndrome scores using the borderline cutoff criterion. Using either of these criteria, a mean of 51.5% of these also had a deviant total problem score. This means that a high total problem score is indicative of a high syndrome score in one half of the children, but

Table 6.4
Mean scores and standard deviations (in parentheses) on CBCL/2-3 scales for boys and girls in the community sample

Scale	Boys (n=215)	Girls (n=205)	Total (N=420)
Oppositional	10.6 (6.0)	10.8 (6.0)	10.7 (6.0)
Withdrawn/Depressed	1.3 (2.1)	1.0 (1.5)	1.1 (1.8)
Aggressive	3.7 (2.9)	2.6 (2.0)	3.2 (2.6)
Anxious	3.3 (3.0)	3.3 (2.8)	3.3 (2.9)
Overactive	3.2 (2.6)	3.0 (2.2)	3.1 (2.4)
Sleep Problems	2.1 (2.6)	2.2 (2.5)	2.2 (2.5)
Somatic Problems	0.3 (0.6)	0.3 (0.7)	0.3 (0.7)
Internalizing	4.5 (4.4)	4.3 (3.6)	4.4 (4.0)
Externalizing	17.5 (9.5)	16.5 (8.8)	17.0 (9.2)
Total Problems	34.4 (17.0)	32.3 (16.6)	33.4 (16.8)

Table 6.5
Variance accounted for by significant effects ($p < .05$) sex, age, and SES on CBCL/2-3 scale scores in the community sample

		Boys (n=215)		Girls (n=205)	
Scale	Sex	Age	SES	Age	SES
Oppositional	–	–	3*	–	–
Withdrawn/Depressed	–	–	–	–	3
Aggressive	5*	–	–	–	5
Anxious	–	2*	–	–	2*
Overactive	–	–	3	–	5
Sleep Problems	–	–	–	–	–
Somatic Problems	–	–	–	3*	–
Internalizing	–	2*	–	–	4
Externalizing	–	–	3	–	2*
Total Problems	–	–	3*	–	3

Note. The significant sex effect indicated higher scores for males; the significant age effects indicated higher scores for older children; the significant SES effects indicated higher scores for children from lower socio-economic status groups. * = Effects most likely to be chance findings.

Table 6.6
Prevalence of one or more deviant syndrome scores, and the proportion of children also having a deviant total problem score

	Above borderline cutoff point			Above clinical cutoff point		
Syndrome	%	*n*	Proportion with deviant total problem score	%	*n*	Proportion with deviant total problem score
Oppositional	5.2	22	90.9	2.4	10	80.0
Withdrawn	3.8	16	75.0	2.1	9	77.8
Aggressive	4.5	19	73.7	2.4	10	50.0
Anxious	4.0	17	58.8	1.9	8	37.5
Overactive	4.0	17	70.6	3.1	13	46.2
Sleep Problems	4.8	20	65.0	2.4	10	50.0
Somatic Problems	5.5	23	47.8	1.2	5	20.0
Internalizing	17.4	73	49.3	10.5	44	50.0
Externalizing	15.7	66	69.7	11.2	47	59.6
Total Problems	15.2	64	--	10.0	42	--
Internalizing and/or Externalizing	27.9	117	51.3	18.3	77	46.8
Any one or more	21.4	90	57.8	12.6	53	45.3

not in the other. The highest correspondence between a deviant syndrome score and a deviant total problem score was found for Oppositional, with an overlap of deviancy of 94.7% and 87.5% using the borderline and clinical cutoff points, respectively.

Table 6.7 shows the distribution of deviant syndrome scores among boys and girls. To compare these proportions, cutoff points were based on the cumulative frequency distributions of problem behaviors pooled across both sexes. Across scales, boys had a higher percentage of deviant scores than girls using either the borderline or the clinical cutoff criterion. The difference between boys' and girls' proportions of deviant scores was significant for the Aggressive scale (borderline: $\chi^2=15.10$, $p=.0001$; clinical: $\chi^2=8.76$, $p=.003$), and for having one or more deviant sydrome scores (borderline: $\chi^2=4.51$, $p=.03$; clinical: $\chi^2=4.61$, $p=.03$). The figures show that virtually all children with very high rates of aggressive behaviors were boys, and that boys had 1.5 to 1.75 times higher rates of deviant scores of any kind. Given the number of analyses, however, these latter two findings might have been obtained by chance alone, using a .05 protection level.

Table 6.7
Proportions of deviant syndrome scores among boys and girls

	Above borderline cutoff point		Above clinical cutoff point	
Syndrome	Boys	Girls	Boys	Girls
Oppositional	5.1	3.9	2.3	1.5
Withdrawn	6.0	3.4	2.3	2.0
Aggressive	8.4	0.5 ***	4.2	0.0 **
Anxious	5.1	2.9	1.9	2.0
Overactive	8.8	4.4	3.7	2.4
Sleep Problems	5.1	4.4	2.3	1.5
Somatic Problems	5.1	5.9	1.4	1.0
Internalizing	13.5	10.7	11.6	9.3
Externalizing	17.2	13.2	12.6	8.3
Total Problems	16.3	14.1	10.7	9.3
Internalizing and/or Externalizing	26.0	20.5	20.5	14.1
Any one or more	28.4	19.5 *	14.4	7.8 *

Note. Cutoff points were based on frequency distributions of problem behaviors pooled across sexes.
* $p<.05$; ** $p<.01$; *** $p<.001$.

6.4 Dutch-American comparison

Correlations between Dutch children's scores on Dutch and American versions of the CBLC/2-3 scales, presented in Chapter 4, indicated strong correspondence between scores on Dutch and American scales for all but one of the scales. For American versions of the five scales for which correlations above .80 were obtained (excluding Somatic Problems), and for the Internalizing, Externalizing, and total problem score we compared the means of scale scores obtained by normative American and Dutch samples of 2-3-year old boys and girls. Scale means and standard deviations for the American normative sample were obtained from Achenbach (1992). The American normative sample was composed of 368 children, equally divided by sex, and randomly selected from a national sample and from communities in the Worcester, Massachusetts metropolitan area. Children who had received mental health services in the preceding 12 months were excluded. (Further details on the sample and the sampling procedure are provided by Achenbach, 1992, and Achenbach et al., 1987). The selection and composition of the Dutch normative sample (209 boys, 201 girls) was described in Chapter 3.

Table 6.8 lists the mean scores for both samples. T-tests showed that for both sexes American children scored significantly higher on the Anxious/Depressed scale ($t=2.07$, $p<.05$, and $t=3.64$, $p<.01$, for nationality differences in boys and girls, respectively)

and the Sleep Problems scale (t=4.36, p<.01, and t=4.05, p<.01,). Given the number of analyses, the nationality difference between boys on Anxious/Depressed may have been obtained by chance using a .01 protection level. We may therefore conclude that, with the possible exception of the Anxious/Depressed and Sleep Problems scales, the magnitudes of scores were similar in normative samples in the two countries. This indicates that a particular score on an American syndrome scale has a similar meaning in the two countries, in that it indicates approximately the same degree of deviance from general population norms (cf. Achenbach et al., 1987).

Table 6.8
Mean scores and standard deviations (in parentheses) among non-referred Dutch and American boys and girls on American CBCL/2-3 scales[a]

	Boys		Girls	
Scale	Dutch (n=209)	American (n=184)	Dutch (n=201)	American (n=184)
Anxious/Depressed	3.7 (2.7)	4.3 (3.0)	4.0 (2.8)	5.1 (3.1)
Withdrawn	3.4 (3.0)	3.5 (2.5)	3.1 (2.6)	3.5 (2.5)
Sleep Problems	2.0 (2.5)	3.2 (2.9)	2.2 (2.5)	3.3 (2.8)
Aggressive Behavior	9.6 (5.1)	8.7 (5.7)	8.6 (4.7)	8.7 (5.7)
Destructive Behavior	4.1 (2.8)	4.4 (3.3)	3.6 (2.3)	4.0 (3.2)
Internalizing	7.1 (5.0)	7.8 (5.5)	7.9 (4.8)	8.6 (5.0)
Externalizing	13.6 (7.0)	13.1 (8.3)	12.2 (6.4)	12.7 (8.2)
Total Problems	33.7 (16.5)	34.0 (20.6)	31.9 (16.4)	34.8 (19.1)

[a] Scales and scores derived from Achenbach (1992).

6.5 Correlates of problem behavior in the community sample

To obtain information on the relationship between environmental factors and problem behaviors in Dutch preschoolers, we investigated the association between demographic, family, and child factors with problem behaviors as assessed by the CBCL/2-3. To the extent that problem behavior as assessed by the CBCL/2-3 is related to these factors in expected directions, these associations also provide evidence of the validity of the CBCL/2-3.

Information obtained from the parent interviews at Time 1 (PI-1) and Time 2 (PI-2) was used to construct single factors to be related to the CBCL/2-3 scale scores. Similar to others (e.g., Rutter et al., 1976; Sameroff et al., 1987), we also composed a set of multi-

ple factors to be related to scale scores apart from the relationship between each single factor and the level of problem behaviors.

The variables that constituted possibly associated factors were derived from PI-1, except the variables on pregnancy and perinatal birth problems, which were derived from PI-2. To enable comparable analyses for all factors, the variables were dichotomized into high- and low-risk categories. For variables with a present-absent reponse format, the dichotomy matched the response format. For variables with more than one response category, the choice of the dichotomy cutpoint was based on significant one-way analyses of variance of CBCL/2-3 scale scores by the variable's categories, with Student-Newman-Keuls post-hoc contrasts. For variables yielding continuous data, the 90th percentile of the cumulative frequency distribution of the responses for that variable was used as a cutpoint.

The description of factors and the percentages of children for which the factor was present are given in Table 6.9. Factors were grouped according to the following broad categories: family composition and social position; child health; parental health; parental psychopathology; family support; maternal parenting attitude; stressful life-events. High-risk levels of each factor were expected to be positively related to each of the CBCL/2-3 scale scores, with two exceptions. Having a mother aged 37 years or older, and having a mother working outside the home for 20 hours per week or more seemed to be protective factors, as suggested by the preliminary analyses of variance.

The relation between these factors and problem behaviors were analyzed by univariate and subsequent multivariate analyses. Differences between mean CBCL/2-3 scale scores, Internalizing, Externalizing, and total problem scores of risk and non-risk groups were first analyzed using *t*-tests. Although analyses of mean scores may be informative as to which factors are associated with a general increase in level of problem scores, in clinical practice it may be of value to identify children at risk for high scores on particular scales. Therefore, the analyses of mean scores was followed by analyses of the proportions of deviant problem scores (above the borderline cutpoint) per factor using χ^2 tests. Factors that were significantly associated with the CBCL/2-3 problem scores are listed in Table 6.10 together with the level of significance of the association according to the *t*-tests (first row following each factor), and χ^2 tests (second row). Asterisks indicate those five significant *t*-test and χ^2 tests effects per syndrome that were most likely to be found by chance, given the number of 68 analyses per problem score, using a .05 protection level (Sakoda et al., 1954).

As can be observed from Table 6.10, 28 of the 34 possibly associated factors showed a significant association with one or more of the problem scores. As may be expected, χ^2 tests showed fewer significant associations than *t*-tests due to the loss of variation in scores using dichotomous scores.

Table 6.9
Sociodemographic, family, and child correlates in the community sample

	Yes		No	
	n	%	*n*	%
Family composition and social position				
Maternal age 37 yrs or higher (-)	59	14.0	361	86.0
First-born	172	41.0	248	59.0
Four or more children in the home	109	26.0	311	74.0
Maternal education less than 10 yrs	235	56.0	185	44.0
Paternal education less than 10 yrs	192	45.7	228	54.3
Family SES unskilled or semiskilled	120	28.6	300	71.4
Father unemployed	31	7.4	389	92.6
Public assistance as portion of family income	35	8.3	379	90.2
Non-european ethnicity	16	3.8	385	3.8
Child health				
One or more pregnancy problems	221	52.6	199	47.4
One or more birth complications	108	25.7	312	74.3
Duration of pregnancy less than 38 weeks	43	10.2	377	89.8
Birth weight less than 2500 grams	22	5.2	398	94.8
Life threatening disease ever	65	15.7	349	84.3
Two or more hospital stays	33	7.9	387	92.1
10 or more days in hospital	41	9.8	379	90.2
One or more serious physical problems ever	49	11.7	371	88.3
Poor health	12	2.9	408	97.1
Parental health				
Poor maternal physical health	33	7.9	387	92.1
Poor paternal physical health	29	6.9	391	93.1
Parental psychopathology				
Parent used mental health service ever	9	2.2	405	96.4
Poor maternal mental health	29	6.9	391	93.1
Poor paternal mental health	22	5.2	398	94.8
Mother impaired by mental health problem ever	33	7.9	380	90.5
Family support				
Single parent	18	4.3	402	95.7
Increase in marital conflicts during last 12 months	8	1.9	406	98.1
Maternal parenting attitude				
30 hours or more nonparental care during first 2 years	39	9.3	381	90.7
20 hours mother out of home	41	9.8	379	90.2
Sometimes or frequently exhausted by raising children	140	33.3	280	66.7
Irritated by child several times per week or more	214	51.0	206	49.0
Physical punishment of child several times per week or more	94	22.4	326	77.6
Self out of control sometimes or often	193	46.0	227	54.0
Battering of child ever	26	6.3	387	92.1
Stressful life-events				
One or more stressful life-events during last 12 months	190	45.9	224	53.3

Table 6.10
Sociodemographic, family, and child correlates associated with problem scores in the community sample

	OPP	WIT	AGG	ANX	OVE	SLE	SOM	INT	EXT	TOT
Family composition and social position										
Maternal age (-)	.000	--	.015	--	.002	--	.001	--	.000	.001
	--	--	--	--	--	--	--	--	--	--
First-born	.002	.027	--	--	.000	.008*	--	--	.002	.002
	--	--	--	--	.000	--	--	--	--	--
Many children	.002	--	--	--	.010*	--	--	--	.003	.002
	--	--	--	--	--	--	--	--	.041*	--
Low maternal education	--	.011	.002	.005	.000	--	--	.001	.003	.001
	--	--	.011*	--	.012	--	--	.011*	.050*	.010
Low paternal education	--	.000	.037*	.003	.000	--	--	.000	.013	.001
	--	.025*	.042*	--	--	--	--	.009*	.017	.001
Low SES	.001	--	.023*	.033*	.000	--	--	.024*	.000	.000
	.017*	--	--	--	--	--	--	--	.001	.001
Father unemployed	.047*	--	--	--	.002	--	--	--	.045*	--
	--	--	--	--	.028*	--	--	--	--	--
Public assistance	--	--	--	--	.001	--	--	--	.038*	--
	--	--	--	--	.010	--	--	--	--	--
Child health										
Pregnancy problems	--	.031*	--	.004	--	--	--	.003	--	--
	--	--	--	--	--	--	--	--	--	--
Birth complications	--	--	--	--	--	.042*	--	--	--	--
	--	.010*	--	--	--	--	--	--	--	--
Short pregnancy	--	--	--	--	--	--	--	--	--	--
	--	--	--	.020*	--	--	--	--	--	--
Life threatening disease	.018	--	--	--	--	.013*	--	--	.020	.003
	--	--	.001	--	--	--	--	--	--	.046*
Hospital stays	.033*	--	.044*	--	--	--	--	--	.019	.004*
	--	--	.028*	--	--	.039*	--	--	--	.046*
Days in hospital	--	.030*	--	--	--	--	--	.021*	.048*	.007*
	--	.019*	--	--	--	.002	--	--	.029*	.029*
Poor health	.000	--	.010	--	--	--	--	--	.011	.000
	.040*	.000	.000	.000	.000	.000	--	.001	.000	.000
Parental health										
Poor maternal physical health	.021*	--	--	.016*	--	.008	.031*	.008	--	.012*
	--	.038*	--	--	--	.004*	.000*	.000	--	.003
Parental psychopathology										
Parental mental health service	--	--	--	.022*	--	--	--	.032*	--	--
	--	--	.010*	.006*	--	--	--	.047*	--	--
Poor maternal mental health	.003	.012	--	.029*	.014*	--	.044*	.009	.002	.002
	.012*	.000	--	.006*	--	--	.004*	.000	--	.000
Poor paternal mental health	--	.008	--	.009	--	--	--	.016*	.026*	.038*
	.001	.000	--	.019*	.026*	--	--	.000	--	.026
Impaired by mental problem	--	--	--	--	--	--	--	--	--	--
	--	--	--	--	--	--	.012*	--	--	--
Family support										
Single parent	.009	--	--	--	.015*	.001	.034*	--	.011	.027
	--	--	--	--	--	.015*	--	--	--	--
Marital conflicts	.019*	--	--	--	--	.001	--	--	--	.015
	--	--	--	--	--	.005*	--	--	--	--
Maternal parenting attitude										
Mother out of home (-)	--	--	--	--	.017*	--	--	--	--	--
	--	--	--	--	--	--	--	.045*	--	--
Exhausted by raising children	.000	--	.009	.004	--	.006	.007	.006	.000	.000
	--	--	--	--	.027*	--	--	--	--	.035*
Irritated by child	.000	.028*	.000	--	.000	--	--	--	.000	.000
	.003*	--	--	--	.008	--	--	--	.000	.000
Physical punishment of child	.005	--	.000	--	.000	--	--	--	.000	.006
	--	--	--	--	.007	--	--	--	.000	--
Out of control	.000	--	.001	--	--	--	.024*	--	.000	.002
	--	--	--	--	--	--	--	--	--	--
Stressful life-events										
Stressful life-events	.000	.032*	.023*	.013	--	.032*	.001	.006	.001	.000
	--	--	--	--	.043*	--	.019*	--	.026*	.004

Note. Rows indicate levels of significance of the association between risk factors and problem scores according to results of t-tests (first row) and results of chi-square tests (second row, bold face). For the chi-square tests, deviant problem scores were defined by a score above T67 for scale scores, and above T60 for Internalizing, Externalizing, and Total problems. Effects most likely to be chance findings are indicated by an asterisk.

Factors having ten or more significant associations with mean problem scores and proportions of deviant scores included low maternal and paternal education, low SES, poor health of the child, poor maternal physical health, poor maternal and paternal mental health, being exhausted by raising children, being irritated by the child, and stressful life-events.

Because the various factors that were univariately related to problem scores are not independent from each other, stepwise logistic regression analyses were performed. Logistic regression is used with dichotomous outcomes to evaluate the predictive value of a set of independent variables. For the independent variables, the strength of association is evaluated using relative odds (the odd's ratio, OR). Relative odds in logistic regression assesses the unique contribution, i.e. separate from the other variables in the model, made by each independent variable to the prediction a deviant problem score. Values greater than 1.0 indicate a positive relation with deviancy; values less than 1.0 indicate a negative relation with deviancy. For each problem score, only those variables were included in the set of independent variables that were significantly related to the problem score in the univariate χ^2 analyses. The SPSSpc logistic regression procedure was used with the statistical criterion for factor entry set at $p < .05$. In Table 6.11 an effect is presented only when the 95% confidence interval around the OR excluded 1.0.

All factors that had many univariate relations with problem scores (except being exhausted by raising children) also showed significant associations in the multivariate analyses.

Of the factors subsumed under the family composition and social position heading, either low SES, low maternal education, or low paternal education showed significant associations with all problem scores except Anxious, Sleep Problems, ans Somatic Problems. These three factors are moderately interrelated (mean Kendall $\tau = .45$), and may to a certain degree be considered as alternative indicators of family social position. Two additional factors under this heading were related to overactive behavior: being first-born, and the family receiving public assistance of any kind.

Of the factors under the child health heading, poor health of the child was related to all problem scores except Oppositional and Somatic Problems. Having had a life threatening disease ever was related to aggressive behavior, and hospitalization of 10 days or more to sleep problems. Of the perinatal factors only birth complications showed an association with withdrawn/depressed behavior.

Maternal physical health was related with the two scales Sleep Problems and Somatic Problems. Poor parental mental health, as indicated by any of the four factors under this heading was related to all scales except Overactive, Sleep Problems, and Externalizing.

Of the family support measures the absence of a father in the home was related to sleep problems.

Maternal parenting attitude measures were exclusively related to externalizing

Table 6.11
Stepwise logistic regression analyses of deviant problem scores in relation to sociodemographic, family, and child correlates

Risk factors	B	SE	*p*	Odds ratio (95% CI)
Oppositional				
Low SES	1.19	0.50	.018	3.3 (1.2-8.7)
Poor paternal mental health	1.75	0.65	.007	5.7 (1.6-20.0)
Irritated	1.70	0.65	.009	5.5 (1.5-19.6)
Withdrawn/Depressed				
Low paternal education	1.07	0.52	.041	2.9 (1.1-8.1)
Poor health	2.17	0.79	.006	8.8 (1.8-42.1)
Poor paternal mental health	2.17	0.63	.000	8.7 (2.7-30.0)
Birth complications	1.34	0.50	.007	3.8 (1.4-10.1)
Aggressive				
Low maternal education	1.44	0.65	.028	4.2 (1.2-15.2)
Poor health	1.87	0.79	.018	6.5 (1.4-30.6)
Life threatening disease	1.30	0.52	.013	3.7 (1.3-10.2)
Parental mental health service	2.20	0.87	.012	9.0 (1.6-49.9)
Anxious				
Poor health	2.36	0.73	.001	10.6 (2.5-44.3)
Parental mental health service	2.21	0.85	.010	9.1 (1.7-48.4)
Overactive				
Low maternal education	1.27	0.50	.011	3.6 (3.6-9.5)
Public assistance	1.10	0.55	.046	3.0 (1.0-8.9)
Firstborn	2.06	0.52	.000	7.8 (2.8-22.0)
Poor health	1.95	0.78	.012	7.1 (1.5-32.1)
Irritated	1.10	0.48	.022	3.0 (1.2-7.6)
Sleep Problems				
Poor health	1.99	0.76	.009	7.3 (1.7-32.5)
Days in hospital	1.54	0.55	.005	4.7 (1.6-13.6)
Single parent	1.56	0.72	.030	4.8 (1.2-19.5)
Poor maternal physical health	1.36	0.59	.021	3.9 (1.2-12.4)
Somatic Problems				
Impaired by mental problem	1.15	0.56	.039	3.2 (1.1-9.4)
Poor maternal physical health	1.47	0.52	.005	4.3 (1.6-12.2)
Internalizing				
Low paternal education	0.76	0.33	.021	2.1 (1.1-11.1)
Poor health	1.53	0.66	.021	4.6 (1.3-16.8)
Poor maternal mental health	1.47	0.49	.003	4.4 (1.7-11.4)
Poor paternal mental health	1.19	0.57	.036	3.3 (1.1-10.1)
Externalizing				
Low SES	0.94	0.30	.002	2.6 (1.4-4.6)
Poor health	1.71	0.61	.006	5.5 (1.6-18.9)
Stressful life-events	0.62	0.29	.033	1.9 (1.1-3.3)
Irritated	1.39	0.33	.000	4.0 (2.1-7.7)
Total Problems				
Low SES	1.03	0.31	.001	2.8 (1.5-5.2)
Poor health	2.32	0.66	.000	10.2 (2.8-37.3)
Poor maternal mental health	1.32	0.46	.004	3.7 (1.5-9.2)
Stressful life-events	0.81	0.31	.009	2.2 (1.2-4.1)
Irritated	1.22	0.34	.000	3.4 (1.7-6.6)

Note. CI = Confidence interval.

behaviors. Being irritated by the child several times a week or more was independently related to oppositional behavior, overactive behavior, and the externalizing and total problem score. Physical punishment of the child several times a week or more was related to overactive behavior.

Finally, stressful life-events occurring during the last 12 months were related only to externalizing behaviors and having a high total problem score.

Summarizing, we may conclude that in the community sample low parental education, low socioeconomic position of the family, poor health of the child, poor parental mental health, maternal parenting attitude, and stressful life-events were the strongest correlates of young preschoolers' problem behavior as assessed by the CBCL/2-3.

Based on the findings from the multivariate analyses we composed a cumulative risk index to analyse the relation between deviant problem scores independent of the specific associated factors. From each group of sociodemographic, child, and family factors the factor with the highest number of significant relations with problem scores was included in the index. The following factors were selected: low SES, poor health, poor maternal physical health, poor maternal mental health, single parent, irritated, and stressful life-events.

The cumulative risk index score in the sample ranged from 0 to 5 (possible range was 0-7). However, the number of children with 4 or 5 factors scored as present was very small (13 and 3, respectively). For purpose of analysis, these children were grouped together with children for whom 3 factors were present, thus yielding four groups with risk index scores of 0, 1, 2, and 3, respectively.

Table 6.12 shows the distributions of deviant problem scale scores over the four risk groups. The strength of the association between group membership and deviant problem score is indicated by (1) the relative risk of having a deviant problem score for children in the group having zero factors scored as present when compared with the group for which 3 factors were present, and by (2) the phi coefficient. Statistical significance was tested by χ^2 tests of independence.

As can be observed from Table 6.12, in terms of relative risk there is evidence for quite strong associations between the number of factors present and having a deviant problem score. The odds ratios of the narrow-band scales are comparable to those obtained for the comparison of referred and nonreferred 2-3-year olds, presented in Chapter 5. Relative risks of deviant problem scores are even higher for the broad-band internalizing and externalizing groupings, and for total problems. Relative risks of having a deviant total problem score was more than 50 times greater for children in the high risk group than for children in the low-risk group. More than 40 percent in the high-risk group had a deviant total problem score. Logistic regression analyses on the scale scores using cumulative risk and sex as independent variables, showed highly significant effects of cumulative risk ($p < .0001$). No significant effects of sex, or risk by sex interactions were found.

Table 6.12
Proportions of children with deviant problem scale scores by cumulative risk

Scale	Cumulative risk 0 (74)	1 (152)	2 (122)	3 (66)	phi	Odds ratio	*p*
Oppositional	1.4	2.0	4.9	15.2	.23	13.21	.000
Withdrawn/Depressed	1.4	3.3	5.7	10.6	.16	8.78	.019
Aggressive	2.7	2.0	7.4	7.6	.13	2.95	.090
Anxious	1.4	3.9	4.1	9.1	.13	7.40	.062
Overactive	2.7	4.6	4.9	19.7	.23	8.83	.000
Sleep Problems	1.4	3.3	4.1	12.1	.16	10.07	.012
Somatic Problems	1.4	3.9	6.6	12.1	.15	10.07	.031
Internalizing	1.4	9.2	16.4	22.7	.21	21.47	.000
Externalizing	2.7	9.9	16.4	39.4	.32	23.40	.000
Total Problems	1.4	8.6	17.2	40.9	.35	50.54	.000
Internalizing and/or Externalizing	4.1	17.1	27.9	51.5	.34	25.15	.000
Any one or more	8.1	17.8	26.2	51.5	.32	12.04	.000

Note. Number of children in each risk group in parentheses. Relative risks were based on the comparison of the risk of deviant problem scores in the group most at risk (4) to the risk in the group least at risk (0). *P*-value obtained from chi-square test of independence.

It should be noted that the applicability of the results from these analyses to the prediction of deviance within general population samples of preschoolers is limited for two reasons. First, the prevalence of deviant problem scores is low, and therefore even high risk ratios indicate relatively few cases in absolute terms. Despite the relatively large differences suggested by the odds ratios, the prediction of individual deviant scores from the cumulative risk index is weak as indicated by the low phi coefficients. The phi coefficients ranged from .13 to .35 (median = .22), suggesting that one can make only weak predictions of the problem scores of indivual children given knowledge of their cumulative risk score. Second, the standard errors around each estimate of associations of factors with deviant problem scores obtained from the logistic regression analyses indicate low levels of accuracy.

6.6 Discussion

Prevalence of problem behavior

A normative data base on the prevalence of problem behaviors can be of value to decide whether a particular child's behavior should be considered deviant or not. Population based data on problem behaviors of 2-3-year-olds as we have presented in this chapter

form a background against which to compare data from other individuals or groups. The prevalence rates of parent-reported problems obtained in this study form a data base that complements the population based information on parent-reported problem behavior in Dutch children aged 4-16, presented by Verhulst, Akkerhuis, and Althaus (1985).

In this chapter certain behaviors, commonly regarded as problematic, were reported by parents to be present in more than fifty percent of the preschool children in a representative population sample. Sixteen items were reported for fifty percent or more of the children in this sample, and may thus be regarded as typical for the age group studied. These reflect the limited emotional and behavioral control, and resistance to socialization, which may be expected in the typical young preschooler. Apparently, each of these items individually reflects behavior which is quite common for this age group. However, also very frequently reported behaviors may be part of a cluster of behaviors, which together reflect a serious problem. Because so many items were scored 1 or 2, the *presence* of those behaviors as such may not indicate a behavioral/emotional problem. For only 18 items, however, 10 percent or more of the children received score 2. For most items a *score 2* may thus more validly indicate that the child's behavior may be regarded as a serious problem. Some items were scored by so few parents, that the children for whom this behavior was reported may be in need of special attention. Twenty-nine items were reported (score 1 or 2) for less than ten percent of the sample. To a large extent, these items reflect highly withdrawn, depressed, and aggressive behaviors, and somatic complaints.

A possibly stronger indication for the presence of a problem is when the individual child's score on an empirically derived syndrome reaches the 95th or 98th percentile of the distribution of scores in the normative population. In this study, 12.4% percent of the 2-3-year olds had one or more syndrome scores above the clinical cutpoint, and 21.6% above the borderline cutpoint.

Age-, sex-, and SES differences

Whereas reports from parents of children aged 4-16 revealed significantly higher total problem scores for boys than for girls, higher scores for younger than for older children, and higher scores for lower than for higher socio-economic status groups, in the present study we found only significant differences related SES at the level of the total problem score.

Looking somewhat closer at the 16 items for which sex differences were found, it appears that all except one item were scored higher for boys than for girls. According to their parents, girls aged 2-3 seem to be more shy than boys. Somewhat higher scores on sleep problems and shyness have also been reported for Dutch girls aged 4-16 (Verhulst, Akkerhuis, and Althaus, 1985). Boys were reported to be consistently more aggressive than girls, as indicated by higher Aggressive scale scores. Based on parent, teacher, and self-reports, in previous studies boys aged 4-18 have been found to score higher than girls on items that could be designated as less socially desirable, such as aggressive and generally disturbing behaviors (Verhulst et al. 1985a, 1985b, 1989). The present study

shows that already at ages 2 and 3 these sex differences can be found. Furthermore, as has been suggested for older children, the qualitative differences between the problem behavior of boys versus girls may account for the higher rates of referral for mental health service for boys than for girls (Eme, 1979; LaClave & Campbell, 1986). Comparable to what has been reported for older children, the ratio of referred boys versus girls in our clinical sample was 2:1 (see Chapter 3). The higher rates of generally disturbing behaviors in boys may lead parents to seek professional help more often for their sons than for their daughters already at the preschool age.

On 13 of the 17 items for which age differences were found, 3-year-olds scored somewhat higher than 2-year-olds. A closer look at these items reveals that higher scores for older children seem to reflect a growing ability to differentiate the self from others. Avoidance of eye-contact, self-consciousness, and withdrawn behavior all reflect that with increasing age children become more and more aware of the presence of other people, who can be readily differentiated from the self. Decreasing resistance to toilet training and the increasing presence of feelings of guilt may reflect that children at this age start to show consideration of other people's wishes. It should be noted that also the first signs of anxious and depressed behavior appear as we know it in older children. Older children in our sample more often were reported to be withdrwan, to worry, and to be too fearful or anxious.

We found small, but significant differences in total problem scores for children from different SES groups. Twenty-two items were scored higher for lower than for higher SES groups. Hyperactive, oppositional, and depressed behaviors were more prevalent in children from lower than higher SES groups. Economic hardship may lead to a diminished capacity for supportive, consistent, and involved parenting (McLoyd, 1990), which may have as negative outcome the higher rates of problems in children, especially at ages, when they are entirely dependent on parental care.

The sex, age, and SES differences in individual problem rates were partially reflected in the scale score differences. The only scale score difference between boys and girls was on Aggressive, on which boys scored higher. Age differences in scale scores were all small, and all could have been found by chance. SES differences were most obvious in girls. It should be noted that all effects of sex, age, and SES on item and scale scores of these young preschoolers were small. As such only the most robust findings might be replicable: boys show more aggressive behaviors, and 2-3-year olds from lower SES show more problem behaviors of several kind than children from higher strata.

Comparison with other studies

The CBCL/2-3 has been used in at least two other studies using large samples. Achenbach (1992) obtained CBCL/2-3's completed by 368 parents of 2-3-year olds in a national American survey sample and children living in Worcester, Massachusetts. With the exception of Anxious/Depressed and Sleep Problems, scores of Dutch and American children on the American (1992) scales were the same. The mean total problem score for

the Dutch and the American sample was 32.8 and 34.4, respectively. Larson, Pless, and Miettinen (1988) using CBCL/2-3's completed by parents of a birth cohort sample of 756 children aged 3 years in Montreal, Canada, obtained a mean total problem score of 32.9, which, again is highly comparable to the mean score obtained in the present study. A comparison of parent ratings of problems for Dutch and American 4- to 16-year olds also revealed no systematic nationality differences (Achenbach et al., 1987).

In our general population sample of preschool children 12.6% had one or more deviant syndrome scores (using the 98th percentile as cutoff point). This compares to 11.1% found in the Canadian study by Larson et al. (1988), who also used the 98th percentile point to obtain prevalence estimates in a birth cohort. Achenbach (1992) reported 12% of the children having one or more deviant scores using the 95th to 97th percentile as cutoff point. Most other studies of behavior problems in preschool children using parental reports and checklists, based the cutoff points on psychiatric judgment. Using the BSQ (Richman & Graham, 1971), most of these studies also estimated the prevalence of severe problem behaviors to be in the range between 6 and 12%. In two studies by Earls (Earls, 1980; Earls & Richman, 1980) larger rates of 24% and 16% were obtained. Since the format of the frequently used BSQ is very different from the format of the CBCL/2-3, it is difficult to directly compare the scores reported for individual items obtained by these instruments.

Correlates of problem behavior

In this community sample, parental education, socioeconomic position of the family, health of the child, parental mental health, and maternal parenting attitude were the strongest correlates of young preschoolers' problem behavior as assessed by the CBCL/2-3. Forty percent of the children for whom three or more selected factors of disadvantage were present had a deviant total problem score. Factors were selectively related to syndrome scores. Internalizing problems were specifically related to low parental education, ill health of the child, poor physical health of the mother, and poor mental health in both parents. Externalizing behaviors were related to low educational and socioeconomic status of the family, ill health of the child, mother being exhausted, irritated, and physical punishing the child, and the presence of stressful life-events. Sleep problems were specifically related to ill health of the child, number of days in hospital, poor maternal health, and marital conflicts. Somatic problems held a specific relation to maternal physical and mental health.

Two other large scale general population studies reported on a wide array of correlates of problem behavior in young preschoolers.

Richman, Stevenson, and Graham (1982) reported that compared to a demographically matched group, 3-year olds with behavior problems came from larger families; received more criticism and hostility, but less warmth from both parents; and had mothers who reported more depression and anxiety, being irritated by the child, physical punishment and loss of control, a poor marital relationship, and a higher number of stressful

events. Unrelated to group status were parental age, length of marriage, work of mother outside the home, birth order, parental physical health, and contacts with family and friends. Our results are highly comparable to the Richman et al. findings. All relations between sociodemographic and family factors found in that study were replicated in ours. Although it was not selected in the multivariate analyses, the presence of four or more children in the home was related to problems also in the present study. Furthermore, unlike Richman's study, our findings also indicated relations between problem scores and maternal age, birth order, and maternal physical health. Unfortunately, the Richman study did not differentiate between different behavior syndromes.

Using the CBCL/2-3 in a birth cohort of 2-3-year olds, Larson, Pless, and Miettinen (1988) found that characteristics associated with most or all problem scales (internalizing, externalizing, sleep problems, somatic problems, and total problems) included ill health in mother, father's lack of help with caretaking, and frequent use of hospital emergency department services. Perceived stress and being firstborn were related to sleep problems. Chronic illness of the child was related to internalizing problems, sleep problems, and total problems. Alternative care of the child was related to total problems. The demographic characteristics maternal education, family income, and SES were related to externalizing and/or total problems. These authors also reported that maternal emotional problems when their child was 18 months old were related to externalizing problems at 2 and 3 years. Although many of the factors used in the Larson et al. study were assessed during the neonatal period, and at 6, 12, and 18 months, and although they were measured in different ways, their findings are in agreement with the findings from the present study. Low SES and maternal education, poor health of child and mother, maternal emotional problems, and (perceived) stress were all among the variables that showed positive relations to problem behavior in our study. Some specific relations, however, were somewhat different among both studies. Our measure of stress (life events) was only marginally related to sleep problems in our study, although stressful events such as hospital visits and marital conflicts were. Unlike Larson's study, alternative care was not positively related to problem behavior. Father's lack of support and income were not directly assessed in our study.

We may conclude that the findings on the relation of important environmental and child factors to CBCL/2-3-scores resulting from this study were in congruence with findings of other epidemiological studies on the same age group. Relations to demographic, child, and family correlates found in other general population studies using both similar and dissimilar indices of problem behavior were replicated to a considerable extent.

Looking at the results from the analyses on associated factors one should bear in mind that the methods used for the collection of the data from which the factors were composed, were less than optimal from a psychometric and methodological point of view. First, items regarded periods of different length. Some regarded the present, others regarded specific periods or events in the past, and some asked for a judgment of the life-

time occurrence of events. These differences between periods covered by the items may have differentially affected the reliability of maternal responses. Second, some items involved present/absent categories, the responses to which were summed into variables. Other involved maternal ratings on ordinal scales. Therefore, the metrics underlying the variables composed from these data were different. Two measures were taken to account for these shortcomings: all data used in the risk analyses were obtained from standardized interviews, and variables were dichotomized to enable identical data-analytic procedures across all data.

A further limitation of these analyses is that the relation between the factors and problem behavior scores was not studied in a longitudinal design. Therefore, based on these findings no conclusions about the direction of the significant relations can be drawn. Most notably, significant associations of ratings of maternal parenting attitude and mental health with the child's problem behaviors may be indicative of bidirectionality. Parental psychopathology and maternal attitudes may have influenced children's behavior. However, child problem behaviors, especially of an externalizing nature, may also have induced certain maternal attitudes and, e.g., feelings of depression. Current theoretical models and empirical findings are indicative of an interactional or transactional relation between these variables rather than a unidirectional one.

Because the analyses were not executed in a longitudinal design even the factors found to be significantly associated with problem scores may not be properly indicated as risk factors. Risk factors were defined by Garmezy (1983) as those factors that, if present, increase the likelihood of a child developing an emotional or behavioral disorder in comparison with a randomly selected child from the general population. These factors must predate the disorder, and must plausibly contribute directly to the occurrence of the disorder. Risk factors may be intrinsic (e.g., biological and genetic attributes of the child) or extrinsic (e.g., family and community factors).

CHAPTER 7

Problem behavior in the clinical sample

In this chapter information is given on the presence of problem behaviors in a sample of 2-3-year- olds referred to outpatient and semi-residential mental health services, and on sociodemographic and clinical correlates of problem behavior in these children. We will provide prevalence rates of behavioral/emotional problems in the sample using the CBCL/2-3; identify differences in prevalence rates related to sex, age, and socioeconomic status; and analyse the association of parent ratings of problem behavior in this sample with sociodemographic information, and with the clinical information on the clients obtained from the mental health services' files.

7.1 Prevalence of individual problem items

The percentages of children in the clinical sample who obtained score 1 and 2 on each specific item are given in the first and second columns of Table 7.1. The prevalence of parent-reported specific problem behaviors in the matched samples of referred and nonreferred children was graphically shown in Appendix E.4.

Individual problem behaviors were reported for quite different numbers of children, ranging from 3.5% for *57. Problems with eyes without medical cause* to 89.3% for *Demands must be met immediately*. The following seven items were scored for less than 10% of the children: *39. Headaches*; *41. Holds breath*; *45. Nausea*; *51. Overweight*; *57. Eye problems*; *78. Stomachaches*; *93. Vomiting*. Six of these items describe physical problems without known medical cause. Thirty-six items were scored for more than 50% of the children. However, only six of these items were scored 2 for 50% or more of the children: *6. Can't sit still*; *8. Can't wait*; *16. Demands must be met*; *59. Quickly shifts activity*; *85. Temper tantrums*; *96. Wants attention*. Each of these items describes behavior indicative of oppositional behavior, and behavioral or emotional undercontrol. Similar behaviors are covered by items that were scored 2 for 40-50% of the children (*5. Can't concentrate*; *15. Defiant*; *20. Disobedient*; *44. Angry moods*; *58. Punishment doesn't change behavior*; *66. Screams*; *81. Stubborn*).

Twenty-five items were scored more often 2 than 1. This indicates that if these items are scored in the clinical sample, they usually reflect frequent or intense behavior in most of the cases. Fifteen of these items have to do with management problems. Others regard

Table 7.1
Frequencies of CBCL/2-3 items scored 1 or 2 and percentage of variance accounted for by significant ($p < .05$) effects of sex, age, and socio-economic status in regression on CBCL/2-3 problem scores in the clinical sample

Item	% Score 1	% Score 2	Sex	Age	SES^
1. Aches	14.2	3.2	3^{F}	2^{O}	2
2. Acts too young	15.5	19.7	--	2^{O}	1^{H}
3. Afraid to try new things	25.8	11.1	--	1^{O}	--
4. Avoids eye contact	28.6	9.0	--	1^{O}	--
5. Can't concentrate	30.3	44.8	--	--	--
6. Can't sit still	29.5	53.5	1^{M*}	--	2
7. Can't stand things moved	20.1	14.2	--	--	--
8. Can't wait	28.2	59.6	--	--	--
9. Chews nonfood	24.7	16.8	--	2^{Y}	--
10. Clings to adults	34.5	23.1	--	--	1
11. Constantly seeks help	38.9	26.0	--	2^{O}	2
12. Constipated	4.8	6.8	1^{F}	2^{O}	--
13. Cries much	32.3	22.7	1^{F}	--	3
14. Cruel to animals	13.5	7.2	1^{M}	--	3
15. Defiant	32.3	47.4	--	$<1^{Y*}$	--
16. Demands must be met	29.9	59.4	--	--	2
17. Destroys own things	30.8	23.1	$<1^{M*}$	--	2
18. Destroys others' things	26.6	20.1	1^{M*}	--	--
19. Diarrhea	14.4	10.9	--	$<1^{Y*}$	--
20. Disobedient	39.5	48.0	--	--	1
21. Disturbed by change	34.1	26.6	--	--	--
22. Doesn't want to sleep alone	17.7	19.4	--	--	2
23. Doesn't answer	41.5	23.1	--	1^{O}	--
24. Does't eat well	30.6	22.9	--	--	--
25. Doesn't get along with other kids	31.4	11.6	1^{M}	--	--
26. No fun	14.0	4.4	--	--	1
27. Lacks guilt	31.2	32.1	--	--	--
28. Doesn't want to go out	8.1	3.3	--	--	--
29. Easily frustrated	32.3	39.5	--	--	1
30. Easily jealous	33.4	32.1	--	--	2
31. Eats or drinks nonfood	10.3	7.0	--	--	1*
32. Fears	27.7	23.6	--	1^{O*}	--
33. Feelings are easily hurt	31.2	25.8	--	1^{O*}	2
34. Accident-prone	25.5	23.6	--	--	3
35. Fights	26.2	15.3	2^{M}	--	3
36. Gets into everything	27.7	31.4	--	--	--
37. Upset by separation	31.9	20.5	--	--	2
38. Can't sleep	21.0	27.3	--	1^{Y}	1*
39. Headaches	3.5	1.1	--	--	<1*
40. Hits others	40.4	25.5	1^{M}	--	<1*

(table continues)

Table 7.1 (continued)

Item	% Score 1	% Score 2	Sex	Age	SES
41. Holds breath	6.3	3.1	1^{F*}	--	--
42. Hurts accidentally	30.3	15.1	--	--	--
43. Looks unhappy	17.7	7.9	--	--	1*
44. Angry moods	33.0	43.9	--	--	3
45. Nausea	6.8	1.7	--	--	--
46. Nervous movements	9.0	13.8	--	--	2
47. Nervous	25.3	13.8	--	$<1^{O*}$	1*
48. Nightmares	24.7	9.6	$<1^{F*}$	--	--
49. Overeating	8.7	6.1	--	--	2
50. Overtired	25.3	11.4	--	--	--
51. Overweight	6.6	0.9	--	--	--
52. Painful bowel movements	6.8	3.3	2^{F}	--	--
53. Attacks people	17.5	9.0	1^{M*}	--	--
54. Picking	21.2	18.3	--	--	1
55. Plays with sex parts	14.8	7.4	--	--	4
56. Clumsy	19.0	14.6	--	--	--
57. Eye problems	2.6	0.9	--	--	--
58. Punishment no behavior change	33.6	45.4	1^{M}	--	--
59. Quickly shifts activity	25.5	60.3	--	--	--
60. Rashes	11.1	9.4	--	--	--
61. Refuses to eat	39.3	16.2	--	--	--
62. Refuses active games	34.9	20.3	--	--	--
63. Rocks head or body	14.2	9.8	--	2^{Y}	1*
64. Resists going to bed	23.8	15.7	--	--	2
65. Resists toilet training	17.9	22.9	--	--	--
66. Screams	34.1	41.3	--	--	3
67. Unresponsive to affection	17.5	7.4	--	--	--
68. Self-conscious	15.7	7.4	--	1^{O*}	2
69. Selfish	27.3	12.4	--	--	--
70. Little affection	23.1	7.0	1^{M}	--	--
71. Little interest	14.6	6.6	--	--	--
72. Little fear	24.7	24.0	--	1^{Y*}	--
73. Shy	31.7	10.7	--	1^{O*}	--
74. Sleeps little	14.2	23.8	1^{F}	2^{Y}	--
75. Smears bowel movements	7.0	4.4	--	2^{Y}	1*
76. Speech problem	12.9	33.6	3^{M}	1^{O}	--
77. Stares blankly	21.4	10.3	--	--	--
78. Stomachaches	7.4	2.2	--	--	--
79. Stores up unneeded things	7.2	3.7	--	--	1
80. Strange behavior	12.7	14.6	1^{F*}	--	1^{H}

(table continues)

Table 7.1 (continued)

Item	% Score 1	% Score 2	Sex	Age	SES
81. Stubborn	32.5	44.5	--	--	--
82. Moody	32.3	32.8	--	--	1*
83. Sulks	34.3	20.3	--	--	1
84. Talks or cries in sleep	26.4	9.2	1^F	--	--
85. Temper tantrums	29.0	50.7	1^{M*}	--	1
86. Too concerned with neat or clean	17.7	8.3	--	--	2
87. Too fearful or anxious	22.5	13.1	--	--	--
88. Uncooperative	48.7	23.4	--	--	--
89. Underactive	9.0	6.6	--	2^O	--
90. Sad	16.6	4.6	--	--	--
91. Too loud	33.2	29.7	--	--	1
92. Upset by new people or situations	26.6	17.5	--	--	--
93. Vomiting	5.9	3.5	--	--	--
94. Wakes often	23.4	30.3	2^F	2^Y	--
95. Wanders away from home	8.1	4.1	--	--	--
96. Wants attention	20.5	67.7	--	--	2
97. Whining	37.1	26.4	--	--	2
98. Withdrawn	21.8	6.3	2^M	--	--
99. Worrying	10.3	4.1	--	1^{O*}	--
100. Other problems	5.7	13.8	--	--	--

Note. N = 458 for multiple regression analyses on problem items. ^{M} = Higher scores for males; ^{F} = Higher scores for females; ^{O} = Higher scores for older children; ^{Y} = Higher scores for younger children; ^{A}: All significant SES effects indicated higher scores for lower SES, except *2. Acts too young*, and *80. Strange behavior* (indicated by superscript) where higher SES scored higher. * = Effects most likely to be chance findings.

overactivity (3), and sleep problems (3). The remaining items (*Acts too young*, *Speech problem*, *Strange behavior*) may be indicative of developmental disorder.

7.2 Sex-, age-, and SES-effects on individual problem scores

To assess the effects of sex, age, and SES on the scores for each problem item in the clinical sample, multiple regressions were performed for all items with sex, age, and SES as independent variables. Table 7.1 displays the percentage of variance in each item score accounted for by significant ($p < .05$) associations with sex, age, and SES. The percent of variance accounted for by each variable is represented by the semipartial R^2 (or part correlation) obtained from the regression analyses after partialling out the effects of any other independent variables that accounted for more variance in the item scores. Using a .05 protection level, 9 out of 100 effects might be significant at $p < .05$ by chance (Sakoda et al., 1954). The effects most likely to be significant by chance are indicated by an asterisk. The items for which significant effects were found are listed in Table 7.2 according to their association with the Internalizing and Externalizing syndrome grouping.

Table 7.2
Items for which significant ($p < .05$) differences between sex, age, and SES effects were found by Internalizing and Externalizing syndrome group in the clinical sample

Effect	Internalizing	Neither	Externalizing
Sex	70. Little affection[M] 80. Strange behavior[F]	1. Aches[F] 12. Constipated[F] 25. Doesn't get along[M] 41. Holds breath[F] 48. Nightmares[F] 52. Painful bowel movements[F] 58. Punishment no change[M] 74. Sleeps little [F] 76. Speech problem[M] 84. Talks or cries in sleep[F] 94. Wakes often[F] 98. Withdrawn[M]	6. Can't sit still[M] 13. Cries much[F] 14. Cruel to animals[M] 17. Destroys own things[M] 18. Destroys others' things[M] 35. Fights[M] 40. Hits[M] 53. Attacks[M] 85. Temper tantrums[M]
Age	2. Acts too young[O] 3. Afraid to try new[O] 4. Avoids eye contact[O] 23. Doesn't answer[O] 68. Self-conscious[O] 73. Shy[O] 89. Underactive[O]	1. Aches[O] 9. Chews nonfood[Y] 12. Constipated[O] 19. Diarrhea[Y] 32. Fears[O] 38. Can't sleep[Y] 47. Nervous[O] 63. Rocks[Y] 72. Little fear[Y] 74. Sleeps little[Y] 75. Smears bowel movements[Y] 76. Speech problem[O] 94. Wakes often[Y] 99. Worrying[O]	11. Constantly seeks help[O] 82. Defiant[Y] 33. Feelings hurt[O]
SES[^]	2. Acts too young[H] 10. Clings to adults 26. No fun 37. Upset by separation 43. Looks unhappy 68. Self-conscious	1. Aches 22. Doesn't want to sleep 31. Eats nonfood 34. Accident prone 38. Can't sleep 39. Headaches 46. Nervous movements 47. Nervous 49. Overeating 54. Picking 55. Plays with sex parts 63. Rocks 64. Resists bed 75. Smears bowel movements 80. Strange behavior[H] 86. Concerned with clean	6. Can't sit still 11. Seeks help 13. Cries much 14. Cruel to animals 16. Demands must be met 17. Destroys own things 20. Disobedient 29. Easily frustrated 30. Easily jealous 33. Feelings are easily hurt 35. Fights 40. Hits 44. Angry moods 66. Screams 82. Moody 85. Temper tantrums 91. Too loud 96. Wants attention 97. Whining

Note. [M] = Higher scores for males; [F] = Higher scores for females; [O] = Higher scores for older children; [Y] = Higher scores for younger children; [^]: All significant SES effects indicated higher scores for lower SES, except *2. Acts too young*, and *80. Strange behavior*, where higher SES scored higher, as indicated by [H].

Sex differences. Sex differences were found on 23 items. Ten of these differences reflected higher scores for girls, and 13 reflected higher scores for boys. All effects were very small, according to Cohen's (1988) criteria. Six sex effects accounted for 2% or more of the variance. Of the 13 items on which boys received significantly higher scores, eight were from the Externalizing syndrome, while only one was from the Internalizing syndrome. The proportion of all externalizing items on which boys scored higher (8 out of 31), however, was not significantly greater than the proportion of all internalizing items on which they scored higher (1 out of 19) (χ^2=2.12, n.s.). Six of the eight effects that indicated higher scores for boys in the Externalizing grouping are included in the Aggressive scale.

Age differences. Age differences were found on 24 items. Nine of these differences reflected higher scores for younger children, and 15 reflected higher scores for older children. All age effects were very small, accounting for 2% or less of the variance. Of the 15 items on which older children received significantly higher scores, two were from the Externalizing syndrome, while seven were from the Internalizing syndrome. The proportion of all internalizing items on which older children scored higher (7 out of 19) was significantly greater than the proportion of all externalizing items on which they scored higher (2 out of 31) (χ^2=5.45, $p<.02$).

SES differences. SES differences were found on 41 items. Each of these differences reflected higher scores for lower SES children, except *2. Acts too young*, and *80. Strange behavior*, where higher SES children scored somewhat higher. All effects were very small. SES accounted for 2% or more of the variance in 23 items. Of the 39 items on which lower SES children received significantly higher scores, 19 were from the Externalizing syndrome, five were from the Internalizing syndrome, while 15 were from neither of these syndromes. The proportion of all externalizing items on which lower SES children scored higher (19 out of 31) was significantly greater than the proportion of all internalizing items on which they scored higher (5 out of 19) (χ^2=4.46, $p<.05$).

7.3 Scale scores in the clinical sample

Tables 7.3 and 7.4 display the scale means and the percentage of variance in each scale accounted for by significant ($p<.05$) associations with sex, age, and SES. The percent of variance accounted for by each variable is represented by the semipartial R^2 (or part correlation) obtained from regression analyses of scales scores on sex, age, and SES. The analyses were performed for the whole sample to detect sex differences, and subsequently for each sex separately to analyse associations with age and SES.

Table 7.3
Mean scores and standard deviations (in parentheses) on CBCL/2-3 scales for boys and girls in the clinical sample

Scale	Boys (n=302)	Girls (n=156)	Total (N=458)
Oppositional	18.0 (8.1)	18.3 (8.7)	18.1 (8.3)
Withdrawn/Depressed	4.1 (3.5)	4.0 (3.5)	4.1 (3.5)
Aggressive	6.9 (4.5)	5.6 (4.3)	6.4 (4.4)
Anxious	5.3 (4.0)	5.4 (3.8)	5.3 (3.9)
Overactive	5.8 (2.7)	5.5 (2.9)	5.7 (2.8)
Sleep Problems	3.9 (3.6)	4.9 (3.9)	4.2 (3.7)
Somatic Problems	0.3 (0.8)	0.6 (1.0)	0.4 (0.9)
Internalizing	9.3 (6.6)	9.4 (6.3)	9.4 (6.5)
Externalizing	30.6 (13.2)	29.5 (13.9)	30.2 (13.5)
Total Problems	61.4 (25.6)	61.5 (26.2)	61.4 (25.8)

Table 7.4
Variance accounted for by significant effects (p<.05) of sex, age, and SES on CBCL/2-3 scale scores in the clinical sample

		Boys (n=302)		Girls (n=156)	
Scale	Sex	Age	SES^{A}	Age	SES^{A}
Oppositional	–	–	2	–	5 *
Withdrawn/Depressed	–	2^{O}	–	–	–
Aggressive	2^{M}	–	–	–	7
Anxious	–	1^{O} *	–	–	–
Overactive	–	–	–	–	–
Sleep Problems	2^{F} *	2^{Y} *	5	–	–
Somatic Problems	2^{F} *	2^{O}	2 *	–	–
Internalizing	–	2^{O}	–	–	–
Externalizing	–	–	2 *	–	6
Total Problems	–	–	3	–	4 *

Note. ^{A}: All significant SES effects indicated higher scores for children from lower socio-economic status groups. ^{M} = Higher scores for males; ^{F} = Higher scores for females; ^{O} = Higher scores for older children; ^{Y} = Higher scores for younger children; * = Effects most likely to be chance findings.

Boys and girls had very similar scale scores and total problem scores. Small sex effects accounting for 2% of variance were found for the Aggressive scale, on which boys scored higher than girls, and for Sleep Problems and Somatic Problems, where girls scored higher than boys. Small age effects (2% of variance or less) were found. Among boys, 3-year olds had higher scores on scales reflecting withdrawn/depressed and anxious behavior, and sleep problems. Two-year old boys were scored higher on Sleep Problems than 3-year old boys. Among girls no significant age effects were found. Significant effects of SES on scale scores reflected higher scores on the Oppositional, Sleep Problems, and Somatic Problems scale, and on Internalizing in boys from lower SES levels. Girls from lower SES received higher scores on Oppositional, Aggressive, Externalizing, and total problems. All effects could be considered small according to Cohen's (1988) criteria. Furthermore, as indicated in Table 7.4, several significant effects could be expected by chance alone, using a .05 protection level (Sakoda et al., 1954).

Like individual items, scale scores may also be used to describe the prevalence of behavioral / emotional problems. This may be presented by the percentage of children having one or more deviant CBCL/2-3 syndrome scores, i.e., scores above borderline or clinical cutoff point. In Table 7.5 these percentages are listed together with the percentage of children who obtained one or more deviant syndrome scores as well as a deviant total problem score. As in previous chapters, the bottom of the borderline range was defined by approximately the 85th percentile on the total problem, Internalizing, and Externalizing score (normalized T-score of 60), and by approximately the 95th percentile on the other scale scores (normalized T-score of 67). The bottom of the clinical range was defined by the 90th percentile on the total problem, Internalizing, and Externalizing score (normalized T-score of 63), and by approximately the 98th percentile on the other scale scores (normalized T-score of 70). A more stringent cutoff point was chosen for the syndrome scores, because the syndrome scores are based on fewer items than the total problem score, implying somewhat less reliability of the first.

In this clinical sample 69.7% of the children had one or more deviant syndrome scores using the borderline cutoff criterion, and 55.7% had one or more deviant syndrome scores using the clinical cutoff criterion. Using either of these criteria, about 85% of these also had a deviant total problem score. This means that a high total problem score is indicative of a high syndrome score in almost seven out of every eight children in this clinical sample. There was a high correspondence between having a specific deviant syndrome score and a deviant total problem score. Proportions of overlap of deviancy on total problems and each of the syndromes exceeded 80% in all cases.

Table 7.5
Prevalence of one or more deviant syndrome scores, and the proportion of children also having a deviant total problem score

	Above borderline cutoff point			Above clinical cutoff point		
Syndrome	%	*n*	Proportion with deviant total problem score	%	*n*	Proportion with deviant total problem score
Oppositional	40.6	186	96.8	26.6	122	96.7
Withdrawn	30.6	140	88.6	21.2	97	89.7
Aggressive	28.2	129	95.3	22.9	105	98.1
Anxious	15.9	73	93.2	12.2	56	89.3
Overactive	21.8	100	94.0	17.2	79	91.1
Sleep Problems	23.8	109	89.9	12.9	59	88.1
Somatic Problems	11.4	52	92.3	4.1	19	89.5
Internalizing	52.4	240	85.0	43.0	197	81.7
Externalizing	59.8	274	93.1	52.8	242	93.4
Total Problems	64.6	296	–	57.6	264	–
Internalizing and/or Externalizing	74.9	343	84.3	66.8	306	83.0
Any one or more	69.7	319	84.6	55.7	255	86.3

Table 7.6
Distribution of deviant syndrome scores among boys and girls in the clinical sample

	Above borderline cutoff point		Above clinical cutoff point	
Syndrome	Boys	Girls	Boys	Girls
Oppositional	37.1	42.9	23.8	30.1
Withdrawn	35.8	35.9	20.2	23.1
Aggressive	32.1	23.7	23.2	16.7
Anxious	15.9	16.0	12.3	12.2
Overactive	31.5	30.8	17.2	17.3
Sleep Problems	20.9	29.5 *	8.6	16.7 **
Somatic Problems	8.9	16.0 *	3.6	5.1
Internalizing	48.3	48.7	42.7	43.6
Externalizing	62.9	59.6	55.0	50.0
Total Problems	63.9	66.0	56.3	59.6
Internalizing and/or Externalizing	74.5	74.4	66.9	68.6
Any one or more	74.8	72.4	52.6	58.3

Note. Cutoff points were based on frequency distributions of problem behaviors pooled across sexes. Significance level of differences between distributions: * $p < .05$; ** $p < .01$.

Table 7.6 shows the distribution of deviant syndrome scores among boys and girls. To compare proportions of deviant syndrome scores for boys and girls, cutoff points were based on the frequency distributions of problem behaviors pooled across sexes. For most scales proportions of deviant scores were very similar for both sexes. Girls had a significantly higher proportion of deviant scores on Sleep Problems (borderline: $\chi^2=4.22$, $p<.05$; clinical: $\chi^2=6.64$, $p<.01$) and Somatic Problems ($\chi^2=4.45$, $p<.05$). In this sample there was only a tendency for boys to have a higher proportion of deviant Aggressive scores than girls ($p=.06$). All of these differences, however, could be chance findings, using a .05 protection level (Sakoda et al., 1954).

7.4 Demographic, child, and family correlates of problem behavior

To obtain information on the relationship between sociodemographic characteristics and problem behaviors in this sample of referred preschoolers, we investigated the association between demographic, child, and family factors with problem behaviors as assessed by the CBCL/2-3. Clinical samples may be highly selective with regard to characteristics of the subject and his demographic and familial background. Background factors associated both with seeking help or coming into contact with mental health services and problem behavior may lead to the finding of spurious relations between problem behavior and demographic, child, and family correlates (Berkson, 1946). However, to the extent that these correlates bear similar relationships to problem behavior in both the clinical sample and the community sample, these may be regarded as robust correlates of problem behavior in young preschoolers.

The information on correlates in this clinical sample was derived from the individual case files available at the participating agencies, as described in Chapter 3. To enable comparable analyses for all correlates, the variables were dichotomized into present versus absent categories. For variables with a present-absent reponse format, the dichotomy matched the response format. For variables with more than one response category, the dichotomy cutpoint matched as exactly as possible the cutpoint chosen in the community sample. Other dichotomy cutpoints were based on significant one-way analyses of variance of CBCL/2-3 scale scores by the variable's categories, with Student-Newman-Keuls post-hoc contrasts.

The description of the correlates and the percentages of children for whom these were coded as present are given in Table 7.7. Asterisks indicate for which correlates information and cutpoints comparable to those used in the community sample were employed. Correlates were grouped according to the following broad categories: family composition and social position; child health; parental psychopathology; family support; parenting history. Parenting history replaced the maternal parenting attitude variable used in the analyses on the community sample, and included ratings of the stability and quality of the caregiving environment during the years preceding intake. Parental health and stressful life-events were not included. Presence of each correlate was expected to be positively

related to each of the CBCL/2-3 scale scores, with the exception of having a mother aged 37 years or older, which was negatively related to some problem scales in the community sample.

Table 7.7
Demographic, child, and family correlates in the clinical sample

	Yes		No	
	n	%	*n*	%
Family composition and social position				
* Maternal age 37 yrs or higher	46	10.0	412	90.0
* First-born	232	50.7	226	49.3
* Four or more children in the home	25	5.5	433	94.5
* Maternal education less than 10 yrs	239	52.2	219	47.8
* Paternal education less than 10 yrs	185	40.4	273	59.6
* Family SES unskilled or semiskilled	191	41.7	267	58.3
* Father unemployed	55	12.0	403	88.0
* Public assistance as portion of family income	42	9.2	416	90.8
* Non-european ethnicity	71	15.5	387	84.5
Child health				
* One or more pregnancy problems	85	18.6	373	81.4
* One or more birth complications	148	32.3	310	67.7
Duration of pregnancy less than 37 weeks	45	9.8	413	90.2
* Birth weight less than 2500 grams	33	7.2	425	92.8
* Two or more hospital stays	56	12.2	402	87.8
Two or more weeks in hospital	86	18.8	372	81.2
Poor health	156	34.1	302	65.9
Parental psychopathology				
Poor maternal mental health	190	41.5	268	58.5
Poor paternal mental health	78	17.0	380	83.0
Family support				
* Single parent	86	18.8	372	81.2
Parenting attitude				
Low stability of caregiving environment	31	6.8	427	93.2
Low quality of caregiving environment	297	64.8	161	35.2

Note. Asterisks indicate those variables that had exactly the same operationalization in the clinical and in the community sample.

The relation between these correlates and problem behaviors was analyzed by univariate and subsequent multivariate analyses. Differences between mean CBCL/2-3 scale scores, Internalizing, Externalizing, and total problem scores of groups with the correlate present or absent were first analyzed using *t*-tests. The analyses of mean scores

was followed by analyses of the proportions of deviant problem scores (above the borderline cutpoint) per correlate using χ^2 tests. Correlates that were significantly associated with the CBCL/2-3 problem scores are listed in Table 7.8, together with the level of significance of the association according to the *t*-tests (first row following each correlate), and χ^2 tests (second row). Asterisks indicate those four significant *t*-test and χ^2 test effects per syndrome that were most likely to be found by chance, given the number of 42 analyses per problem score, using a .05 protection level (Sakoda et al., 1954).

As can be observed from Table 7.8, 20 of the 21 correlates showed a significant association with one or more of the problem scores.

Because the various correlates that were univariately related to problem scores are not independent from each other, stepwise logistic regression analyses were performed. For the independent variables, the strength of association was evaluated using the odd's ratio (OR). For each problem score only those variables were included in the set of independent variables that were significantly related to the problem score in the univariate χ^2 analyses. The SPSSpc logistic regression procedure was used with the statistical criterion for factor entry set at $p < .05$. In Table 7.9 an effect is presented only when the 95% confidence interval around the OR excluded 1.0.

Fourteen of the 20 correlates that had relations with problem scores in the univariate analyses also showed significant, independent associations in the multivariate analyses.

Of the social position variables, low SES, unemployment of the father, and receipt of public assistance as part of the family income did not enter any of the logistic models. Maternal age was positively related to high Anxious scores. Being firstborn related negatively to internalizing behavior, i.e., second or later born children had higher odds on having a deviant Internalizing score. Having four or more children in the home was related to deviant Aggressive as well as Anxious and Internalizing scores. Low paternal education significantly contributed to the models for aggressive as well as overactive behavior, whereas low maternal education was independently related to deviant Internalizing, Externalizing, and total problem scores. Having one or more parents of noneuropean origin was associated with high Internalizing scores.

Of the group of child health variables, pregnancy problems, low birthweight, and poor health had no independent contribution in the analyses. The presence of birth complications was related to high levels of anxiety. Having had a pregnancy duration of less than 37 weeks contributed to the model for externalizing behavior. Two or more hospital stays during the preceding years was related to Overactive and Total Problems. Having stayed in hospital during a total of two weeks or more contributed to the models for internalizing as well as externalizing problems, i.e., deviant Oppositional, Withdrawn/Depressed, Aggressive, Anxious, and Internalizing scores.

Table 7.8
Demographic, child, and family correlates significantly associated with problem scores in the clinical sample

	OPP	WIT	AGG	ANX	OVE	SLE	SOM	INT	EXT	TOT
Family composition and social position										
Maternal age	--	--	--	--	--	--	--	--	--	--
	--	--	--	.016	--	--	--	--	--	--
First-born (-)	--	--	--	.006	--	--	--	.018	--	--
	--	--	--	--	--	--	--	.011	--	--
Many children	--	--	.011	.000	--	--	--	.000	--	.003
	.042*	--	.023	.024	--	--	--	.001	--	--
Low maternal education	.006	--	.025	--	.010	--	--	--	.003	.001
	.037*	--	.026*	--	.026*	--	--	.010	.004	.002
Low paternal education	.036*	--	.012	--	.002	--	--	--	.006	.022*
	--	--	.021	--	.007	--	--	--	--	--
Low SES	.004	--	.023	--	--	.011	--	--	.006	.002
	.016*	--	.032*	--	--	--	.029*	--	.014*	.002
Father unemployed	--	--	--	.049	--	.015	--	.037	--	.028*
	--	--	.016	--	--	.020*	--	.039	--	--
Public assistance	--	--	--	--	--	.046*	--	--	--	--
	--	--	.026*	--	--	--	--	--	--	--
Noneuropean ethnicity	--	.019*	--	.000	--	.011	--	.002	--	.009
	--	.041*	--	--	--	.031*	--	.005	--	--
Child health										
Pregnancy problems	--	--	--	--	--	--	.030*	.028	--	--
	--	--	--	.034	--	--	--	--	.046*	--
Birth complications	--	--	--	--	--	--	--	--	--	--
	--	--	.038*	.004	--	--	--	--	.019*	--
Short pregnancy	--	--	--	--	--	--	--	--	--	--
	--	--	--	--	--	--	--	--	.004	--
Low birthweight	--	--	--	--	--	--	--	--	--	--
	--	--	--	--	.036*	--	--	--	--	--
Hospital stays	--	.015*	.004	.019	.001	--	--	.006	.005	.002
	--	.033*	.009	--	.002	--	--	.006	.029*	.008
Weeks in hospital	.012	.003	.007	.000	.042*	--	--	.000	.004	.000
	.007	.011	.002	.001	.003	--	--	.002	.005	.035*
Poor health	--	--	--	--	--	--	--	.045*	--	.049*
	--	--	--	--	--	--	--	.043*	--	.021
Parental psychopathology										
Poor maternal mental health	.000	--	.002	.008	--	.024*	--	--	.000	.000
	.001	--	.001	--	--	--	--	.010	.000	.000
Family support										
Single parent	--	--	--	.025	--	.003	.020	--	--	--
	--	--	--	--	--	.007	.048*	--	--	--
Parenting history										
Low stability of care	.000	.001	--	.000	--	--	--	.000	.003	.001
	--	--	--	.000	.018*	--	--	.000	.001	.002
Low quality of care	.000	--	.000	.021	.001	.005	.016	.041*	.000	.000
	.000	--	.004	--	.008	.009	.025*	.042*	.000	.000

Note. Rows indicate levels of significance of the association between risk factors and problem scores according to results of t-tests (first row), and results of chi-square tests (second row, bold face). For the chi-square tests, deviant problem scores were defined by a score above T67 for scale scores, and above T60 for Internalizing, Externalizing, and Total Problems. Effects most likely to be chance findings are indicated by an asterisk.

Table 7.9
Stepwise logistic regression analyses of deviant problem scores in relation to demographic, child, and family correlates

Scale / Correlates	B	SE	*p*	Odds ratio (95% CI)
Oppositional				
Weeks in hospital	0.59	0.25	.018	1.8 (1.1-2.9)
Quality of care	1.10	0.22	.000	3.0 (1.9-4.6)
Withdrawn/Depressed				
Weeks in hospital	0.62	0.25	.012	1.9 (1.1-3.0)
Aggressive				
Many children	0.86	0.43	.044	2.4 (1.0-5.4)
Low paternal education	0.47	0.22	.030	1.6 (1.0-2.4)
Weeks in hospital	0.79	0.26	.002	2.2 (1.3-3.6)
Poor maternal mental health	0.74	0.22	.000	2.1 (1.4-3.2)
Anxious				
Maternal age	0.99	0.38	.009	2.7 (1.3-5.6)
Many children	1.04	0.47	.027	2.8 (1.1-7.1)
Birth complications	0.72	0.29	.011	2.1 (1.2-3.6)
Weeks in hospital	0.64	0.31	.040	1.9 (1.0-3.5)
Stability of care	1.50	0.42	.000	4.5 (2.0-10.1)
Overactive				
Low paternal education	0.57	0.23	.014	1.8 (1.1-2.8)
Hospital stays	0.89	0.31	.004	2.4 (1.3-4.5)
Quality of care	0.62	0.26	.018	1.9 (1.1-3.1)
Sleep Problems				
Single parent	0.58	0.26	.028	1.8 (1.1-3.0)
Quality of care	0.55	0.25	.029	1.7 (1.1-2.8)
Somatic Problems				
Quality of care	0.78	0.35	.028	2.2 (1.1-4.4)
Internalizing				
Firstborn	- 0.43	0.21	.035	0.6 (1.0-0.4)
Many children	1.39	0.57	.015	4.0 (1.8-12.3)
Low maternal education	0.40	0.20	.044	1.5 (1.0-2.2)
Noneuropean ethnicity	0.61	0.29	.034	1.8 (1.1-3.2)
Weeks in hospital	0.71	0.26	.008	2.0 (1.2-3.4)
Poor maternal mental health	0.44	0.20	.033	1.5 (1.0-2.3)
Stability of care	1.40	0.51	.006	4.1 (1.5-10.8)
Externalizing				
Low maternal education	0.51	0.20	.012	1.7 (1.1-2.5)
Short pregnancy	1.23	0.40	.002	3.4 (1.6-7.5)
Stability of care	1.22	0.55	.028	3.4 (1.1-10.0)
Quality of care	0.97	0.21	.000	2.6 (1.7-4.0)
Total Problems				
Low maternal education	0.52	0.20	.011	1.7 (1.1-2.5)
Hospital stays	0.86	0.37	.019	2.4 (1.2-4.8)
Stability of care	1.32	0.63	.036	3.7 (1.1-12.8)
Quality of care	0.77	0.21	.000	2.2 (1.4-3.3)

Note. CI = Confidence interval.

Of the parental psychopathology correlates only poor maternal mental health had an independent contribution to Aggressive and Internalizing scores. Paternal mental health did not enter any model.

The only significant contribution of being raised by a single parent was to Sleep problems. Children living with one parent had a relatively large chance of having a deviant Sleep Problems score.

Both low stability and low quality of the caregiving environment were related to various problem scales. Low stability was related to deviant Anxious, Internalizing, Externalizing, and total problem scores. A history characterized by a low quality of the caregiving environment was related to deviant Oppositional, Overactive, Sleep problems, Somatic problems, Externalizing, and total problem scores.

Summarizing, we may conclude that in this sample of young preschoolers referred to various mental health services the number of children in the home, the level of parental education, frequency and duration of hospital stays, maternal mental health, and stability and quality of care were the strongest correlates of problem behavior as assessed by the CBCL/2-3.

7.5 Differences between children referred to different types of mental health service

The children included in this clinical sample were recruited from mental health services with different levels of specialization, and serving different populations of children: one child guidance consultation agency; four outpatient community mental health centers; four child psychiatric outpatient clinics, three of which were university clinics and one operated by an independent hospital; and three day clinics for children up to age six who are maladjusted or who show developmental problems. The services may be described along the following general outlines.

The child guidance consultation agency (in Dutch: Pedagogisch Consultatie Bureau; PCB) was known to be consulted by parents referred by general practitioners, and by pediatricians of general consultation agencies, a service that is provided free of charge in the Netherlands for children below age 4. Children were mainly referred for behavioral problems or suspected developmental delay. If children were judged to be in need of specialized help, they were referred to outpatient settings or other specialized services. The agency was closed in 1992 for financial reasons.

The four child departments of the outpatient community mental health centers (in Dutch: Regionale Instelling voor de Ambulante Geestelijke Gezondheidszorg; RIAGG) are all located in the city of Rotterdam,and serve densely populated urban areas and suburban areas with a total population of about 1 million people. Most children are referred to the RIAGG by the general practitioner or by schools. According to year reports published by the centers, about 25% of the parents contact the RIAGG without referral.

The child psychiatric outpatient clinics, three of which are part of a university hospital, generally serve large geographic regions. They provide highly specialized diagnostic, consultation, and treatment services to children referred by medical specialists,

RIAGGs, and other medical, psychological, school, and justice institutions.

Of the three Medical Daycare Centers for Preschool Children (in Dutch: Medische Kleuter Dagverblijven; MKD) two were located in the city of Rotterdam, and one in the eastern part of the Netherlands. The MKDs in the Netherlands have a total capacity of 1774 places. Children between 0 and 6 years (mean age 3.9 yrs) are placed in the centers because somatic, psychological, and social factors place them at serious risk for developmental disorders and retardation. According to several sources (Twijnstra, Hüsken, De Jonge, & Olthoff, 1992; Van Yperen, 1990) about 70% of the MKD population has a psychopathological disorder according to adapted ICD-criteria, 50-80% has one or more developmental disorders, more than 30% has a subnormal level of intelligence, 60-80% has one or more abnormal somatic conditions, and more than 90% is living in abnormal social conditions. Most children are referred to the MKD by medical specialists, the RIAGGs, and the general practitioner. At average children are placed in the centers for 15.8 months.

These four types of services may be ordered according to the model of the referral pathway to psychiatric care for children presented by Verhulst and Koot (1992). This model describes five levels of psychiatric morbidity, including four levels of referral. The levels of psychiatric morbidity represent circumscribed populations at the level of the community, primary health care, conspicuous primary care morbidity (the level representing children identified by the primary care physician as disturbed), specialist psychiatric care, and child psychiatric inpatient care. The child population at each level can be seen as a sample of the population at a lower level. To move from one level to the next, the child has to pass through a "filter" describing the selection processes that determine for which children treatment will be sought at a certain level. In this model, the child guidance agency may be placed at the conspicuous primary health care level, the community mental health centers and child psychiatry units at the specialist psychiatric care level, and the medical daycare centers at the highest, (semi)residential care level.

Representing different levels of referral and specialization of care, the mental health institutions described here may be expected to serve children and parents referred by different sources, who need help for problems of different type, duration, and/or intensity. We analyzed differences between the four types of services with regard to referral source of the clients, services previously received by the clients, referral question, referral complaints, and level of problem behavior reported by the parents. Because ICD-10 diagnoses had to be based on diagnostic file information and services differed greatly in the completeness of information available, a comparison of diagnoses across service types was omitted.

Table 7.10
Proportions of children from different referral sources by mental health service

Referral source	Child Guidance (n=56)	Community MH (n=172)	Child Psychiatry (n=135)	Daycare Centers (n=91)	Total
Parents	1.8	8.1	1.5	3.3	4.4
Preschool settings	7.1	2.9	4.4	3.3	4.0
Primary health care	53.6	47.7	36.3	25.3	40.5
Specialist health care	1.8	4.7	37.0	34.1	19.8
Mental health professionals	0.0	3.5	10.4	28.6	10.1
Other	35.7	33.1	10.4	5.5	21.1

Referral source. As shown in Table 7.10, the majority of referrals were made by primary (40.5%) and specialist health care (19.8%). A minor proportion of children were referred by parents and preschool settings (about 4% each). About 10% of the referrals was made by mental health professionals, in most cases to services at a higher level of specialization. Twenty-one percent of the referrals came from various other professionals and institutions. The higher a service was situated at the referral pathway, the more children had been referred by high-level sources (χ^2=159.26, p=.000; Mantel-Haenszel test for linear association = 16.13, p=.000).

Referral question. Services differed considerably with regard to the main referral questions posed at intake. The majority of questions posed to the child guidance consultation agency regarded parent guidance (91.1%), whereas all referral questions to medical daycare centers regarded placement of the preschooler. Children were referred to community mental health centers mostly for diagnostic examination (70.3%) or parent guidance (23.4%). The majority of referrals to child psychiatric outpatient care regarded diagnostic (81.6%) or developmental (8.8%) examination.

Previous help. Of all children in the sample, 32.5% had received previous help from one other service, and 10.0% had received help from two or more services. The proportions of children that had received help previously were 5.4, 36.6, 50.0, and 65.9, for the four types of services, respectively, which is a significant difference (χ^2=59.7, p=.000; Mantel-Haenszel test for linear association = 40.49, p=.000).

Referral complaints. Referral complaints were obtained from the referral letter and/or from the intake report. For all children except three, one or more complaints were coded (mean = 2.17). The majority of the children presented with conduct problems (56%), developmental problems (35%), and/or management problems (27%). Emotional, sleep,

Table 7.11
Proportions of referral complaints by mental health service

Referral complaint	Number of children	Percentage of children	Percentage of complaints	Child Guidance (*n*=56)	Community MH (*n*=175)	Child Psychiatry (*n*=136)	Daycare Centers (*n*=91)	Chi-square	p
Developmental problems	159	34.7	18.1	7.1	28.0	39.7	57.1	43.9	.000
Emotional problems	57	12.4	6.4	3.6	11.4	16.2	14.3	6.2	.101
Conduct problems	257	56.1	30.6	71.4	51.4	52.2	61.5	8.8	.032
Temper tantrums	38	8.3	3.8	19.6	2.9	9.6	9.9	16.9	.000
Sleep problems	65	14.2	6.5	19.6	17.1	11.8	8.8	5.5	.141
Contact problems	56	12.2	8.0	0.0	9.7	21.3	11.0	19.4	.000
Management problems	125	27.3	12.8	50.0	32.6	11.0	27.5	35.1	.000
Family problems	59	12.9	6.6	0.0	13.1	8.1	27.5	28.3	.000
Eating problems	27	5.9	2.7	10.7	4.6	6.6	4.4	3.4	.335
Toileting problems	10	2.2	1.3	3.6	2.3	2.2	10.0	1.0	.797
Other	27	5.9	2.7	1.8	9.1	3.7	5.5	6.3	.099

Table 7.12
Mean scores and standard deviations (in parentheses) on CBCL/2-3 scales for children referred to different mental health services

Scale	Child Guidance (*n*=56)	Community MH (*n*=175)	Child Psychiatry (*n*=136)	Daycare Centers (*n*=91)	F-ratio	p	Contrast
Oppositional	20.7 (6.1)	17.8 (8.4)	16.2 (8.5)	20.0 (8.2)	6.10	.0004	2 < 4; 3 < 1,4
Withdrawn/Depressed	3.7 (3.3)	3.6 (3.1)	4.0 (3.4)	5.3 (4.1)	5.31	.001	1,2,3 < 4
Aggressive	7.5 (3.7)	6.3 (4.4)	5.5 (4.3)	7.5 (4.7)	5.21	.001	3 < 1,4
Anxious	4.9 (3.7)	4.8 (3.8)	5.3 (4.3)	6.4 (3.6)	3.61	.01	2,3 < 4
Overactive	6.2 (2.4)	5.4 (2.9)	5.2 (2.9)	6.6 (2.3)	6.19	.0004	2 < 1,4; 3 < 4
Sleep Problems	5.3 (3.3)	4.0 (3.7)	4.0 (3.9)	4.4 (3.7)	2.09	.10	--
Somatic Problems	0.3 (0.7)	0.4 (0.8)	0.4 (0.9)	0.6 (1.0)	1.46	.22	--
Internalizing	8.6 (6.4)	8.4 (5.9)	9.3 (6.6)	11.7 (7.0)	5.61	.001	1,2,3 < 4
Externalizing	34.4 (9.9)	29.4 (13.8)	26.9 (13.6)	34.2 (13.0)	7.65	.0001	2,3 < 1,4
Total Problems	66.4 (20.9)	58.1 (25.3)	58.1 (25.3)	71.3 (27.3)	7.67	.0001	2 < 1,4; 3 < 4

Note: 1 = Child Guidance Consultation Agency; 2 = Community Mental Health Centers; 3 = Child Psychiatric Outpatient Clinics; 4 = Daycare Centers.

contact, and family problems were each presented by 12-14% of the children. Temper tantrums, eating problems, and toileting problems were each presented by less than 10% of the children. The proportion of complaints by service type is shown in Table 7.11. The distribution of complaints across services types differed for six of the 11 complaint groups as tested by the χ^2-test. The proportion of developmental problems reported as complaint at intake ranged from 7.1% for the child guidance center to 57.1% for the daycare services, and was linearly related to the level of specialization of services ($p=.000$). Conduct problems were reported for more than half of the children in all types of services, but somewhat more in the child guidance service (71.4%) and daycare centers (61.5%) than in the outpatient and child psychiatric settings. Temper tantrums were reported for relatively many children referred to the child guidance service. The child psychiatry departments received more of the children with reported contact problems than any other service. The proportions of children for whom management problems were reported were also very different for the various types of services, ranging from 11.0% at the child psychiatry units to 50.0% at the child guidance service, and was inversely related to level of service specialization ($p=.000$). Finally, family problems were reported much more for the daycare centers (27.5%) than for any of the other services.

Problem scores. To see whether the levels of parent-reported problem behaviors differed among children from different types of services, a MANOVA was performed on the CBCL/2-3 problem scale scores of the children from the four different service groups, followed by oneway ANOVAs with post-hoc Student-Newman-Keuls contrasts. The Internalizing and Externalizing scales were not included in the MANOVA because these are linearly dependent on the narrow-band scales. Instead, separate ANOVAs were performed on these scales. The MANOVA was highly significant (Pillai's trace = .109; $p=.001$). Table 7.12 gives the mean scale scores and the results from the univariate analyses. Significant differences were found for all scales except Sleep Problems and Somatic Problems. Children referred to the medical daycare centers had higher scores than the other groups on the Withdrawn/Depressed, Internalizing, and Anxious scales, although the difference with the child guidance agency was not significant for this latter scale. Further, clients at the medical daycare centers had higher scores than those referred to community mental health and child psychiatry units on Oppositional, Aggressive, Overactive, Externalizing, and total problem scales. Children seen at the child guidance agency received higher scores on the Overactive, Externalizing, and total problem scale than those referred to community mental health settings, and higher scores on Oppositional, Aggressive, and Externalizing than clients of child psychiatry units.

Since these differences in problem scores may be attributable to demographic differences among children referred to these services, we analyzed whether the distributions of the children's sex, age, and SES were different for these services. Only SES was differentially distributed across the services ($\chi^2=19.26$, $p<.004$). Whereas the distribution of children across low, medium, and high SES levels was similar for the community

Table 7.13
CBCL/2-3 scale scores significantly associated with referral complaints

Scale	Developmental (*n*=159)	Emotional (*n*=57)	Conduct (*n*=257)	Temper tantrums (*n*=38)	Sleep (*n*=65)	Contact (*n*=59)	Management (*n*=125)	Eating (*n*=27)
Oppositional	↓ .000	--	↑ .000	↑ .001	--	↓ .001	↑ .000	--
Withdrawn/Depressed	↑ .000	--	--	--	↓ .004	↑ .000	--	--
Aggressive	↓ .001	↓ .001	↑ .000	--	--	↓ .015	↑ .000	--
Anxious	--	↑ .000	--	--	--	--	↓ .015	--
Overactive	--	--	↑ .000	--	--	--	↑ .000	--
Sleep Problems	↓ .000	↑ .004	↑ .008	--	↑ .000	↓ .002	--	--
Somatic Problems	--	↑ .008	--	--	--	--	--	↑ .000
Internalizing	↑ .001	↑ .000	--	--	↓ .019	↑ .008	--	--
Externalizing	↓ .000	--	↑ .000	↑ .009	--	↓ .001	↑ .000	--
Total Problems	--	--	↑ .000	--	--	--	↑ .000	--

Note. Entries indicate whether the scores of the complaint group were significantly higher (↑) or lower (↓) than those of the noncomplaint group as indicated by univariate *F*-tests, and the significance level of the *F*-tests.

Table 7.14
Mean CBCL/2-3 scale scores by ICD-10 Axis I diagnosis

Scale	No disorder (*n*=276)	Hyperkinetic (*n*=41)	Conduct (*n*=54)	Mixed Cond/Emot (*n*=18)	Emotional (*n*=36)	Social function. (*n*=14)	Other (*n*=19)	*F*-ratio	*p*	Contrast
Oppositional	16.6	20.0	22.6	22.2	18.9	20.8	15.2	6.31	.000	1 < 2,3,4; 7 < 3
Withdrawn/Depressed	3.8	4.2	4.3	5.3	4.4	6.2	3.6	1.86	.087	--
Aggressive	5.6	9.2	9.1	8.5	5.5	7.6	3.8	10.84	.000	1,7 < 2,3,4; 5 < 2,3
Anxious	4.8	5.6	5.2	5.4	7.9	7.0	5.3	3.81	.001	1,2,3 < 5
Overactive	5.4	7.0	6.6	6.2	5.5	6.5	4.1	4.93	.000	1,7 < 2,3
Sleep Problems	4.1	3.4	5.1	4.9	4.3	4.4	3.5	1.16	.338	--
Somatic Problems	0.4	0.2	0.6	0.4	0.4	0.9	0.5	1.26	.276	--
Internalizing	8.6	9.9	9.5	10.7	12.3	13.2	8.9	2.97	.007	1 < 5
Externalizing	27.6	36.2	38.3	36.9	29.9	34.9	23.2	9.15	.000	1,7 < 2,3,4; 5< 3; 7 < 6
Total Problems	56.9	70.2	73.2	71.5	64.6	79.9	50.0	6.21	.000	1,7 < 2,3

Note. 1 = No disorder; 2 = Hyperkinetic disorders; 3 = Conduct disorders; 4 = Mixed disorders of conduct and emotions; 5 = Emotional disorders; 6 = Disorders of social functioning; 7 = Other behavioral and emotional disorders.

mental health centers and child guidance agency (approximately 5:3:2), the medical daycare centers had more clients from lower strata (6.5:2.5:1). The child psychiatry units had equal proportions from each stratum (1:1:1). Inclusion of SES as a covariate in the MANOVA still yielded significant differences (Pillai's trace = .105; p=.008). Although still significant, the differences between services on the Oppositional, Aggressive, Overactive, and Externalizing scales were somewhat smaller when corrected for SES.

7.6 Clinical correlates of CBCL/2-3 scores

Referral complaints. Children were grouped according to whether or not a specific complaint was coded from the files. To analyze the association between referral complaints and levels of parent-reported problem behavior MANOVAs were performed on the cluster of CBCL/2-3 problem scale and total problem scores with complaint (present vs. not present) as a grouping variable. If the overall MANOVA was significant, univariate F-tests were carried out for individual scales. The Internalizing and Externalizing scales were not included in the MANOVAs because these are linearly dependent on the narrow-band scales. Instead, separate ANOVAs were performed on these broad-band scales.

All CBCL/2-3 scales were differentially related to referral complaints ($p < .01$ or less) (see Table 7.13). Children referred because of developmental problems received higher Withdrawn and Internalizing scale scores, and lower Oppositional, Aggressive, Sleep Problems, and Externalizing scores than children without developmental problems at intake. Children having emotional problems had higher Anxious, Sleep Problems, Somatic Problems, and Internalizing scores, and lower Aggressive scores than other children. Children seen for reasons of conduct had higher scores on the Oppositional, Aggressive, Overactive, and Externalizing scales, on Sleep Problems and on total problems. Children referred with temper tantrums only had higher Oppositional and Externalizing scores. Children who had sleep problems as referral complaint had higher Sleep Problems scores, and lower Withdrawn and Internalizing (p=.019) scores. Children who were seen because of contact problems had higher Withdrawn and Internalizing, but lower Oppositional, Sleep Problems, and Externalizing scores. They also tended to have significantly lower Aggressive scores (p = .015). Children who's behavior constituted a management problem for their parents had higher Oppositional, Aggressive, Overactive, Externalizing, and total problem scores, and tended to have significantly lower Anxious scores (p = .015). Children with eating problems were scored significantly higher only on Somatic Problems. Children for whom family relation problems were reported as referral coplaints did not differ significantly on any of the scales from children for whom this complaint was not coded. The group of children referred because of toileting (n = 10) contained too few subjects to perform MANOVAs on their scale scores.

Because considerable proportions of the complaints subsumed under the broad headings given in Table 7.13 could not be further differentiated, it was not feasible to do more fine-grained analyses of the relation between referral complaints and parent reports

of problem behavior. For example, although 24% of the complaints subsumed under the conduct problems heading indicated hyperactivity and 19% indicated aggression, 47% of the complaints under this heading were not further specified.

ICD-10 diagnoses. ICD-10 categories that were used in the present analyses regarded behavioral and emotional disorders with onset specific to childhood or adolescence (categories F90-F98, denoted as Axis I), disorders of psychological development (categories F80-F89, denoted as Axis II), and mental retardation (categories F70-F79, denoted as Axis III). Axis IV categories covering somatic diseases were used as a child correlate in the analyses reported in section 7.4 above. Axis V ratings of psychosocial conditions were not used in the present analyses because their reliability could not be computed.

Almost two-fifth of the children in the clinical sample (39.7%, n = 182) received an ICD-10 Axis I diagnosis (categories F90-F98; Behavioral and Emotional Disorders), the majority of which were Hyperkinetic disorders (22.5%), Conduct disorders (29.7%), and Emotional disorders (19.8%). Forty-six percent of the hyperkinetic disorders were specified as disturbances of activity and attention, and 29.3% as Hyperkinetic conduct disorder. Other subcategories regarded hyperkinesis or were unspecified. Seventy-eight percent of the children who received the diagnosis of conduct disorder were placed in the Oppositional defiant disorder category. Other subcategories were Conduct disorder confined to the family context (11%), and Unsocialized conduct disorder (3.7%). Specified emotional disorders regarded Separation anxiety disorder (16.7%), Social sensitivity disorder (25%), or Sibling rivalry disorder (5.6%). Of the social functioning disorders, 64.2% were specified as Reactive attachment disorder, one was specified as Elective mutism, and one as Disinhibited attachment disorder. Specified other behavioral and emotional disorders regarded enuresis, encopresis, feeding disorder, pica, and stuttering.

To see whether the levels of parent-reported problem behaviors differed among children assigned to different diagnostic categories, one-way ANOVAs were performed on the CBCL/2-3 problem scale scores of the children from the different diagnostic groups, followed by Student-Newman-Keuls contrasts. As can be seen in Table 7.14, significant differences were found on seven of the 10 scales. Oppositional behavior scores were lower for the No disorder group than for the Hyperkinetic, Conduct, and mixed conduct and emotional groups. The group with Other diagnoses had lower Oppositional scores than the Conduct disorder group. Although mean Withdrawn/Depressed scores differed clearly among the groups, with the highest scores for children diagnosed with Social functioning disorder, these differences did not reach overall significance. Aggressive scores were significantly higher for the hyperkinetic, conduct, and mixed groups than for the No disorder and Other disorders groups. Children diagnosed as Emotionally disordered had lower aggressive scores than Hyperkinetic and Conduct disordered

children. Scores on the Anxious scale were higher for the Emotional disorder group than for the hyperkinetic, conduct, and mixed disorder groups. Overactive scores were significantly higher in the Hyperkinetic and Conduct disorder groups than in the groups with no disorder or other disorders.

As may be expected, significant differences were also apparent on the broad-band and total problem scales. Emotionally disordered children had higher Internalizing scores than children without a diagnosis. Children diagnosed with Hyperkinetic, Conduct, or mixed Conduct and Emotional disorder had significantly higher externalizing scores than children without diagnosis or Other disorder. Furthermore, higher externalizing scores were found in Conduct disordered than in Emotionally disordered children, and in children diagnosed with Social functioning disorder than in children with the diagnosis Other disorder. Finally, the Hyperkinetic and Conduct disorder groups had significantly higher total problem scores than the No disorder and Other disorder groups. Although the group of Social functioning disordered children had the highest mean score on total problems, differences with other diagnostic groups did not reach significance.

ICD-10 Axis II diagnoses (categories F80-F89, Developmental Disorders) were ascribed to 38.6% (n = 173) of the children in this clinical sample (see Table 7.15). These were specified as developmental disorders of speech and language (37.5%), developmental disorder of motor function (4.0%), Mixed developmental speech/language and motor function disorder (29.5%), Pervasive developmental disorder (19.1%), and Developmental disorder not otherwise specified (9.8%). To analyse differences between developmental disorder groups on parent-reported problem behaviors again one-way ANOVAs were performed on the CBCL/2-3 problem scale scores of the children from the different diagnostic groups, followed by Student-Newman-Keuls contrasts ($p < .05$). We found significant differences between the developmental disorder groups on six of the 10 scales ($p = .000$, except for total problems: $p = .02$). Oppositional scores were higher for the No disorder group than for the Language, Mixed, and Pervasive disorder groups, and for the Language disorder than for the Pervasive developmental disorder group. Children diagnosed with Pervasive developmental disorder had higher Withdrawn/Depressed scores than children without a diagnosis. Aggressive scores were higher for the No disorder, Language disorder and Mixed disorder groups than for the Pervasive developmental disorder group. Sleep problem scores were higher for the No disorder group than for the Mixed and Pervasive developmental and the Other developmental disorder groups, and for the Language disorder than for the Pervasive developmental disorder group. Externalizing scores were higher in the No disorder, Language and Mixed disorder groups than in the Pervasive developmental disorder group, and in the No disorder than in the Mixed disorder group. Finally, total problem scores were highest in the group of children who received no developmental disorder diagnosis, which was significantly higher than in the

Table 7.15
Mean CBCL/2-3 scale scores by ICD-10 Axis II diagnosis

Scale	No disorder (*n*=281)	Speech/ language (*n*=65)	Mixed (*n*=51)	Pervasive (*n*=33)	DD-NOS (*n*=17)	*F*-ratio	*p*	Contrast
Oppositional	19.7	17.5	15.8	11.8	15.2	9.76	.000	2,3,4 < 1; 4 < 2
Withdrawn/Depressed	3.6	4.6	4.3	6.2	4.8	5.37	.000	1 < 4
Aggressive	6.9	6.6	5.9	3.4	5.9	5.22	.000	4 < 1,2,3
Anxious	5.3	5.8	5.4	5.0	4.6	0.43	.788	--
Overactive	5.8	5.6	6.0	4.7	4.9	1.70	.148	--
Sleep Problems	4.9	4.0	3.0	2.0	2.4	8.22	.000	3,4,5 < 1; 4 < 2
Somatic Problems	0.5	0.4	0.3	0.2	0.2	1.48	.206	--
Internalizing	8.9	10.4	9.7	11.2	9.4	1.49	.203	--
Externalizing	32.4	29.6	27.6	19.9	26.1	8.23	.000	4 < 1,2,3; 3 < 1
Total Problems	63.7	62.1	58.8	49.8	52.8	2.96	.020	4 < 1

Note. Speech/language = Speech/language disorders; Mixed = Mixed speech/language and motor function disorder; Pervasive = Pervasive developmental disorder; DD-NOS = Developmental disorder not otherwise specified.

Pervasive developmental disorder group.

Intelligence and mental retardation were scored on a scale having eight categories ranging from *1. very high intelligence* to *7. profound mental retardation*, and *8. mental retardation unspecified*. The mental retardation scales were conform the ICD-10 F70-F79 (Mental Retardation) categories. To analyse the relation between CBCL/2-3 scale scores of children of different levels of intelligence, children in categories 1 (very high intelligence) through 3 (average intelligence) were taken together to form one group, children with low intelligence (IQ 70-85) formed a second group, and children in categories 5 (mild mental retardation) through 8 (mental retardation unspecified) formed a third group. Oneway ANOVAs on the scale scores of these groups were performed. However, because in five of the 10 analyses the homogeneity of variances differed significantly (Bartlett's *F*-test, $p<.01$), groups 2 and 3 were taken together to form a low intelligence/mental retardation group ($n=89$) to be compared with the normal/high intelligence group ($n=231$). The normal/high intelligence group scored higher than the low intelligence/mental retardation group on the Oppositional ($F=12.86$, $p=.000$), Sleep Problems ($F=10.24$, $p=.001$), and Externalizing scale ($F=8.16$, $p=.005$), but lower on the Withdrawn/Depressed scale ($F=6.99$, $p=.009$).

7.7 Discussion

In this chapter a study on a sample of 458 2-3-year-old boys and girls referred to various mental health settings was reported. Parents or caretakers completed the Child Behavior Checklist for Ages 2-3 at intake, and demographic, child, family, and diagnostic information was obtained from clinical case files.

Compared to same-age children in the general population, CBCL problem items were scored for high proportions of children and were more often scored as *2* (Very true or often true of the child). Eleven of the 28 items that were more often scored *2* than *1* in this clinical sample were among the 16 items scored for more than 50% of the community sample. Thus, high scores on these items may partly reflect exaggerations of age-specific problem behaviors for a considerable proportion of the clinical sample. All of these were part of the externalizing grouping, denoting oppositional and overactive behavior. The other items scored more often *2* than *1* regarded sleep problems, developmental problems, uncorrectable behavior, lack of frustration tolerance, nervousness, and toilet problems. Apart from the items denoting developmental problems, all of these high-frequency items imply management difficulties that may have caused parents to seek professional help.

As was the case in the community sample, most items on which boys were scored higher than girls describe aggressive behavior. In the clinical sample, however, girls had higher scores than boys on items that describe sleep and toileting problems. These differences were also reflected in higher aggressive scale scores in boys, and higher scores on the sleep problems and somatic problems scales in girls. Thus, even among

referred children boys show somewhat more aggressive behavior than girls. However, again all sex effects were small. Furthermore, proportions of deviant scale scores were not consistently different for the two sex groups. Apparently, these data cannot account for the 2:1 sex difference in referral rates found in this study, and also reported by others (e.g., Eme, 1979). If parents report not consistently more problem behavior for boys than for girls, then why is the number of referred boys twice the number of referred girls? One explanation may be that the type of problems seen in boys may raise more concern in adults than the type of problems seen in girls, and thus lead to more help seeking behavior. Two findings are supportive of this suggestion. In the commmunity sample virtually all children with very high aggressive scores were boys. Further, in the clinical sample a larger proportion of boys than girls (39.4% vs. 25.6%) presented with developmental problems at intake. However, these differences in proportions alone cannot account fully for the difference in referral rates. A second, and and additional explanation is provided by the finding that in the community sample the proportion of boys with one or more deviant syndrome scores was twice the proportion of girls.

From these findings it may be concluded that once children at this age are referred for professional help, levels and types of problem behavior may be expected to be not dramatically different for boys and girls.

Age differences in the levels of individual problem items showed only partial overlap with age differences found in the community sample. Of the 23 items on which differences were found in the clinical sample only seven also occurred in the community sample. Consistent across samples was the finding that internalizing problems increase with age. Age effects at the level of scale scores indicated higher levels of internalizing (withdrawn and anxious behavior) and somatic problems in older boys, and more sleep problems in younger boys. For girls, no age effects at the scale score level were found. In the community sample, older boys also had higher levels of internalizing problems (anxiety), whereas girls had more sleep problems. Thus, the finding of more internalizing problems in older children is only consistent across samples for boys.

In this sample, SES differences in reported problem behaviors were found for more than 40% of all items, i.e., for twice as many items as in the community sample. Half of these differences regard items denoting externalizing behaviors. Virtually all differences indicated higher scores for children from lower than from higher SES levels, which is consistent with the finding in the community sample. In the clinical sample, high levels of externalizing behavior (oppositional and aggressive) were most prominent in girls from lower SES groups. Apparently, if girls from lower SES groups are referred for professional help they are likely to show more acting out problems than girls from other SES levels, whereas this is somewhat less the case with referred boys from the same socioeconomic stratum. Two items, *Acts too young* and *Strange behavior*, were reported more in higher SES groups than in lower SES groups. This may reflect the higher proportion of children diagnosed as Pervasive developmental disorder in higher (18.8%) than in middle

and lower (both 4.7%) socioeconomic status levels.

In summary, mean CBCL/2-3 scale scores were in this referred sample on the average almost two times as high as in the community sample. The small sex, age, and SES effects on scale scores were barely comparable to those obtained in the community sample. Almost 70% of the children had one or more deviant syndrome scores using borderline range cutpoints. Unlike in the community sample, deviant scale scores concurred with deviant total problem scores in at least 80% of all cases. Using the same cutpoint for both sexes, girls had a somewhat higher proportion of deviant sleep problems and somatic problems scores. However, the large difference between boys and girls on proportion of deviant Aggressive scores was not present in this sample. As noted above, levels of parent reported acting out behaviors are not different for girls and boys, once they are referred.

In multivariate analyses of this referred sample, each of the behavioral/emotional syndromes was related to at least one and at most seven of the demographic, child, and family factors. The number of children in the home, the level of parental education, frequency and duration of hospital stays, maternal mental health, and stability and quality of care were the strongest correlates of problem behavior as assessed by the CBCL/2-3. Low maternal education, ill health as indicated by hospitalization of the child, and caretaking history were separately or in combination prominent correlates of all problem scales. Caretaking history was differentially related to internalizing and externalizing problems. Internalizing problems (most notably anxious behavior) were only related to the stability of care experienced by the child. Narrow-band scales reflecting externalizing problems were exclusively related to the quality of care. Several factors were only related to specific scales and apparently had no significant independent contribution to either the group of externalizing or internalizing scales. Examples are maternal age, birth order, ethnicity, maternal mental health, the number of children in the home, and single parenthood.

Comparison of the findings on correlates of problem behavior in the clinical and community sample shows that indices of educational/occupational level of the parents, indices of child health or hospitalization, parental mental health, caregiving, and single parenting were related to the same problem scales in both samples. Of the narrow-band externalizing scales Oppositional was related to (quality) of caregiving, Aggressive to parental education, child health and parental mental health, and Overactive to parental education, child health and (quality) of care in both samples. Of the narrow-band internalizing scales both Withdrawn and Anxious were only related to indices of poor health or hospitalization. The only correlate of Sleep Problems recurring in both samples was single parenthood. For somatic problems no recurring correlates were found. Consistent correlates of the broad-band Internalizing scale were parental education, child health and parental mental health. Both Externalizing and total problem scores showed consistent relations to parental education, child health and caregiving history.

As discussed in Chapter 6, these findings largely confirm those found in the few other reported community studies on the same age group using parent reports of problem behavior. One remarkable finding of our study was that across samples only indices of child health and hospitalization were consistent correlates of withdrawn and anxious behavior. This finding is consistent with those reported by others on the relationship between problem behaviors in young preschoolers and, for example, chronic illness (Larson, Pless, & Miettinen, 1988), hospitalization (Thompson, 1985), or the sequelae of very low-birthweight (Weisglas-Kuperus, Koot, Baerts, Fetter, & Sauer, 1993).

Most children in this sample were referred to mental health services by primary and specialist health care. The majority of the sample presented with conduct problems (56.1%), developmental problems (34.7%), and/or management problems (27.3%). Based on the clinical file information present at the services almost 40% of the children received a psychiatric diagnosis, the majority of which were hyperkinetic (23%), conduct (30%) and emotional (20%) disorders.

Problem behaviors as reported by the parents were significantly related to type of mental health service. For children referred to semi-residential settings higher levels of internalizing problems were reported than for children referred to other services. These daycare centers, but also the child guidance consultation agency provided help to children with more externalizing problems than those referred to community mental health and child psychiatry units. Referral information showed that the four types of services differed in predictable ways with regard to the referral source of the clients, the number of services previously received by the clients, the type of referral question, and the type of referral complaints. These data confirm the stepwise ordering of the level of specialization of the services along the referral pathway as proposed above. To the extent that being referred to a more specialized level mental health service implies more serious psychopathology in terms of type, frequency, and duration, one may expect to find differences between services in levels of problem behavior on various syndromes reported by parents of children who came for help. Therefore, the differences between types and levels of problem behavior among children referred to various services found in this study may be regarded as evidence of the discriminant validity of CBCL/2-3 syndrome scores.

CBCL/2-3 scales were also differentially and in predictable directions related to referral complaints and ICD-10 diagnoses based on file information. This is an important finding, because it relates parent reports on problem behavior to information on problems obtained from independent sources. File information was scored by raters blind to the CBCL/2-3 scores. Only few of the parents (4.4%) came directly to the services without intervention of a professional or institution. As a consequence, in most cases referral complaints obtained from the file information reflect reasons for referral as formulated by other informants such as primary health care physicians, medical specialists, and mental health workers. The associations between complaints and diagnoses and syndrome scores found in this study therefore further corroborate the construct validity of the CBCL/2-3 syndromes.

The results obtained in the present study are only partially comparable to those obtained in other reported studies on samples of referred children at preschool age. Hooks, Mayes, and Volkmar (1988) reported that in a child psychiatric sample about one-third of the children presented with developmental delay, one-third with behavior problems, and one-third with emotional problems and abuse/neglect. In our subsample of children referred to outpatient child psychiatry units these proportions were .40, .52, and .16, reflecting that considerably more children presented with behavior problems and fewer with emotional problems in our sample. The overall percentage of 56% of children presenting with conduct problems, however, compares well with the 58% reported by Lee (1987) on a sample referred to a military psychiatry unit. Rescorla (1986), and Beitchman, Weckerle, and Hood (1987) studying children referred to the Yale Child Study Center and a psychiatric preschool program, respectively, reported on proportions of children receiving DSM-III diagnoses. However, these figures are not directly comparable to the proportions of diagnoses found in the present sample, because of the diagnostic system used in the present study, the range of mental health services studied, and the material on which diagnoses were based.

Some of the conclusions from this study that are based on information obtained from the clinical case files should be interpreted with caution. Although the information was coded using standardized methods, the interrater reliability of some of the variables was only moderate, i.e., kappas were below .60. Disagreement between raters was in many cases due to the differential use of the category "unknown". This was most notably the case when ratings had to be based on incomplete information. Nevertheless, where findings appeared to replicate across samples, or when information obtained from independent sources showed relations in expected directions conclusions may be regarded as valid.

CHAPTER 8

Predictive value of parent ratings of young preschoolers' problem behavior

Two years after the first time of measurement, the community sample was approached again in a prospective follow-up study. The two times of measurement will be designated as Time 1 and Time 2, respectively.

The aims of the study were:

1. to test the stability and change of parents' reports of children's problem behavior between Time 1 and Time 2;
2. to determine how well parents' reports of problem behavior at Time 1 predict teacher reports of problem behavior at Time 2;
3. to determine how well family and child variables at Time 1 and 2, parents' reports of children's problem behavior at Time 1, and parents' reports of intervening stressful experiences, child temperament, parenting stress, and teacher ratings of children's competence predict parent and teacher ratings at Time 2;
4. to identify predictors of Time 2 parent and teacher reports of significant problems; and
5. to identify correlates of persistence and change of parents' ratings of preschoolers problem behavior.

8.1 Sample and procedure

Parents who participated in the study at Time 1 were contacted between November 1991 and June 1992. A letter was sent to the parents of the 420 children explaining the purpose of the follow-up and indicating that an interviewer would contact them by telephone to make an appointment for an interview analogous to the interview at Time 1 (see Appendix H.1). The parents were contacted by telephone and visited by one of four trained female interviewers, who had an education at the master's level in psychology or special education, and were blind to the Time 1 data.

Usable parent information was obtained for 397 of the 420 children participating at Time 1 (204 boys, 193 girls) with a mean age of 5.31 years (SD=0.64). The respondents were 98.0% mothers, 1.8% fathers, and 0.3% other. Ethnic background was 95.7% Caucasian, 1.6% Surinam/Antillean, 0.3% Turkish, and 2.5% other. Addresses were

obtained for all children except three, who had moved (two of whom abroad). The response rate corrected for untraceable children was 95.2% of the Time 1 sample, or 86.7% of the original target sample. The mean interval between the Time 1 and Time 2 assessment was 2.27 years (SD=0.24).

During the home visit the interviewer read the Child Behavior Checklist (CBCL/4-18) problem items regarding the target child aloud, and scored the parent's responses. After completing the CBCL/4-18, the parent was asked questions about family constellation, health, behavioral outcome, the need of professional help, and the use of services. At the end of the interview, the Language Screening Instrument (LSI), the Dimensions of Temperament Survey-Revised (DOTS-R), the Nijmegen Parenting Stress Index (NPSI), and the Life Events Questionnaire (LEQ) were explained and handed over to be filled out by the parent, together with a postpaid envelope to return these questionnaires. Further, parents were asked for permission to have the child's teacher complete several checklists, including the Teacher's Report Form (TRF). After the visit, parents were promised to be sent a concise general report of the study. A concise report concerning the development of the individual child based on the data obtained with the LSI, DOTS-R, NPSI, and LEQ was sent to the parents one month after the end of the data collection.

Teachers of children whose parents gave permission to address them were sent the problem part of the TRF, the Nijmegen Observation Scales for Preschoolers (NOSP), the LSI, and a questionnaire including questions on significant behavioral/emotional problems of the child, the need of professional help, and the use of mental health and school services, together with an introductory letter, and a prepaid envelope (see Appendix H.2, for the introductory letter). Usable teacher information was obtained for 83.1% (n=349) of the 420 children participating at Time 1, or 76.2% of the original target sample.

Nonresponders

Because Time 1 information was available on those children on whom information was lacking at Time 2, it was possible to trace differences between responders (n=397) and nonresponders (n=23). Dropouts and remainers had almost identical Time 1 total problem scores on the CBCL/2-3 (mean = 31.6 vs. 33.5, t = 0.41, n.s.). The groups also were not significantly different on any of the CBCL/2-3 scales, sex, age, SES, and maternal and paternal education. The only significant difference between both groups was that a larger proportion of dropouts came from non-Caucasian than from Caucasian ethnic groups (18.2% vs. 4.8%; Fisher's exact test p = .026). Although children from non-Caucasian ethnic groups were somewhat more likely to dropout from the study, initial deviance on problem behavior was not associated with selective attrition.

8.2 Instruments

The instruments used in the Time 2 assessment were described in detail in Chapter 3. The scoring of the instruments that were included in the predictive analyses is described below.

The stability and persistence of ratings of preschoolers' problem behavior were analyzed using correlations and odds ratios. To assess the contribution of Time 1 parent ratings in the prediction of Time 2 parent and teacher ratings of problem behavior and significant problems, path analyses were performed using four sets of predictor variables. These predictor sets included (1) Time 1 family and child variables; (2) Time 1 parent ratings of problem behavior; (3) parent reports of stressful life-events that occurred between Time 1 and Time 2; and (4) Time 2 family and child variables.

Predictors

Time 1 Disadvantage

To cover Time 1 family and child characteristics, the disadvantage score constructed from environmental, family and child correlates of CBCL/2-3 scores was used. This score included the presence of one or more of the following conditions: low SES, poor child health, poor maternal physical health, poor maternal mental health, single parenthood, maternal irritation, and one or more stressful life-events preceding Time 1. The construction of the Disadvantage score has been described in Chapter 6.

Time 1 General Development

As an index of the Time 1 developmental level of the child, the General Development scale score derived from the Minnesota Child Development Inventory was used. The construction of this predictor variable was described in Chapters 3.

Time 1 parent ratings

Time 1 parent ratings on the CBCL/2-3 were scored for the narrow-band, broad-band, and total problem scales as described in Chapter 4.

Stressful life-events

To assess life-events that are likely to induce stress in the family or individual, parents filled out a questionnaire containing the 32 items from the Life-Events Questionnaire developed by Berden (1992). They were asked to indicate whether an event had occurred during the 12 months preceding the Time 2 interview. One open item was added to score life-events that were not mentioned in the questionnaire. Only items were included that expressed or implied a negative event, and items that were not spuriously related with problem behavior scores (see section 3.3.2). The items had a yes/no format to indicate whether or not an event had occurred during the 12 months before the Time 2 assessment.

Usable LEQ data were obtained from 369 parents. The item scores (0, 1) were summed into a total score. Events mentioned in the open item were scored present if they met the inclusion criteria, and were added to the total event score.

Time 2 family and child variables

Time 2 family and child correlates of parent and teacher rated problems were derived from the Time 2 Parent Interview, and from standardized scales described below.

Family correlates included (a) Time 2 SES; (b) marital status of the parents other than married or widowed; (c) number of siblings in the home; and (d) mental health services received by family members since Time 1. From the Parent Interview items covering child health, a 4-point parent rating of the child's health during the preceding 12 months was used (1 = excellent; 4 = poor).

Language Screening Instrument (LSI)

To assess language problems at preschool age, we used the Language Screening Instrument (LSI; Gerritsen, 1988), consisting of a parent and a teacher rating, The LSI parent rating was completed by 368 parents and the teacher rating by 348 teachers. Using the ratings as a screening instrument, as suggested by the author of the instrument, a parent score was computed by summing the scores of the four most discriminating items (cf. Gerritsen, 1988), and in the same way a teacher score. Items covered the following aspects of language use: comprehensibility; fluency; word finding; pronunciation; grammar; understanding. The two scores had each a possible range of 4 to 12. Cronbach's alpha internal consistencies was .71 for both the parent and teacher ratings, which is close to the .74 reported by Gerritsen (1988) for the parent ratings, but somewhat below the .84 reported for the teacher ratings. A language problems score was computed by summing the scores of these two parts of the instrument. High scores indicate more language problems.

Nijmegen Observation Scales for Preschoolers (NOSP)

The Nijmegen Observation Scales for Preschoolers (NOSP; Rost, 1992) was to tap preschoolers' level of school-related competence using teacher ratings. The NOSP-scales were completed for 348 children. Scales were scored by summing the scores of items belonging to each scale. Cronbach's alpha internal consistencies in our sample were .89, .91, .77, and .66 for the Task-related behavior, Social behavior, Affect, and Self-help scale, respectively. Although the internal consistency of the Self-help scale was somewhat low, no items could be deleted to raise its reliability. A total School Competence score was computed by summing the four scale scores after transformation to z-scores. High scores indicate high School Competence.

Dimensions of Temperament Survey-Revised (DOTS-R)

To assess preschoolers' difficult temperament according to parent ratings, the Dimensions of Temperament Survey (DOTS-R; Winlde & Lerner, 1986) was used. DOTS-R questionnaires were completed by 373 parents. For each of the DOTS-R scales, a score was computed by summing the raw scores of each scale after recoding reversed items. The following Cronbach's alpha consitencies were obtained for the nine scales: Activity Level-General .75, Activity Level-Sleep: .84 (after deletion of item 46, which reduced

alpha to .73); Approach-Withdrawal: .80; Flexibility-Rigidity: .73; Mood: .72; Rhythmicity-Sleep: .55; Rhythmicity-Eating: .81 (after deletion of item 46, which reduced alpha to .75); Rhythmicity-Daily Habits: .54; and Task-Orientation: .71. The internal consistencies obtained in the present sample were all lower than those reported by Windle and Lerner (1986) for American preschoolers, but were well within the range of alpha's reported by the same authors for a sample of late elementary school children. Although the internal consistencies for Rhythmicity-Sleep and Rhythmicity-Daily Habits were low, these scales were included in the scoring to achieve comparability of results across studies.

To construct a temperament score, the procedure proposed by Windle (In press) was followed. Windle (In press) constructed a difficult temperament index on the basis of the difficult temperament construct of Thomas and Chess (1977), that varied in score range from 0 to 6. If a child received a score of six on the derived difficult temperament index, it indicated a profile of arrhythmicity, withdrawal from novel stimuli, inflexibility, negative mood quality, high activity level, and low task orientation. The derived difficult temperament construct is consistent with, but not identical to the Thomas and Chess constellation. Using this and a similar difficult temperament index, Windle (1991; In press) found that in adolescents the number of difficult temperament factors was associated with more childhood behavior problems (scored retrospectively), and with concurrent depressive symptoms and delinquency, as well as with reduced levels of perceived family and friend support. We constructed the same difficult temperament index as Windle (In press) did by deriving a dichotomous (0, 1) score for each of the eight temperament attributes, with a "1" indicative of a score below the 30th percentile for seven of the attributes, and above the 70th percentile for the attribute activity level-general. Scores of "1" were assigned if any of the three rhythmicity dimensions were below the 30th percentile, indicative of arrhythmicity, making the maximum score for arrhythmicity 1. Further, scores of "1" were assigned each to approach-withdrawal, flexibility-rigidity, mood quality, task-orientation, and activity level-general if they met the respective cutoff criterion. Activity-sleep was not considered in the derivation of the index, because it was not part of the Thomas and Chess difficult temperament constellation.

Nijmegen Parenting Stress Index (NPSI)

To assess the level of perceived parental stress originating from several child and parent characteristics within the caregiving context, we used the Nijmegen Parenting Stress Index (NPSI; De Brock, Vermulst, & Gerris, 1990), which is an adaptation of Abidin's (1983) Parenting Stress Index. Due to a printing failure which was corrected after two weeks of data collection, usable NPSI data were obtained from only 352 parents. Because items in the NPSI assess parent as well as child domains, the NPSI score might be spuriously related to the child's problem behavior. Therefore, an *NPSI-Parent* (NPSI-P) score was computed by summing only the scores on the items reflecting perceived parent characteristics (11 items). Example items are *I have much more problems with raising children than I had expected*, and *I often feel that I can't manage things that well*. High scores

indicate high perceived parenting stress. The internal consistency of the NPSI-P (Cronbach's alpha) in this sample was .85.

Outcomes

Child Behavior Checklist for Ages 4-18 (CBCL/4-18)

The Child Behavior Checklist for Ages 4-18 (CBCL/4-18; Achenbach, 1991b) was used to obtain parent ratings of behavioral/emotional problems of the child. CBCLs were completed by 397 parents in the follow-up sample. For each of the eight CBCL/4-18 cross-informant syndrome construct scales, and for the Internalizing and Externalizing groupings, a score was computed by summing all scores 1 and 2 for the items they comprise. A total problem score was computed by summing all the 1 and 2 scores for all items (except *2. Allergy* and *4. Asthma*, because they do not discriminate well between referred and nonreferred children; see Achenbach, 1991b). If the parent had rated more than one problem item for *113. Other problems*, only one problem receiving the highest score was counted toward the total problem score.

Teacher's Report Form (TRF)

We used the Teacher's Report Form (TRF; Achenbach, 1991c) to obtain teacher ratings of behavioral/emotional problems of the child. TRFs were completed for 349 children in the follow-up sample. For each of the eight TRF cross-informant syndrome construct scales, and for the Internalizing and Externalizing groupings, a score was computed by summing all scores 1 and 2 for the items they comprise. A total problem score was computed by summing all the 1 and 2 scores for all items. If the teacher had rated more than one problem item for *113. Other problems*, only one problem receiving the highest score was counted toward the total problem score.

Parent and teacher ratings of significant problems

Using information from the Parent Interview and the Teacher Questionnaire obtained at Time 2, six signs of significant problems were scored as present versus absent over the period between the Time 1 and Time 2 assessment. These variables indicated whether the child (a) was judged by the parent to have a significant behavioral/emotional problem; (b) was judged by the teacher to have a significant behavioral/emotional problem; (c) was reported by the parent to have received mental health services or special help for academic problems in school; (d) was reported by the teacher to have received mental health services or special help for academic problems in school; (e) was judged by the parent to need professional help for behavioral/emotional problems, that was not being received; (f) was judged by the teacher to need professional help for behavioral/emotional problems, that was not being received. The six signs of significant problems were summed into a total disturbance score with a possible range of 0 to 6.

8.3 Representativeness of Time 2 CBCL/4-18 and TRF data

To assess the similarity of the CBCL/4-18 and TRF data obtained in the community sample at Time 2 (the 1991 sample) to what might be expected in Dutch 4-5-year olds, we compared them to the Dutch normative data obtained in 1983 by Verhulst (Verhulst, Akkerhuis, & Althaus, 1985; Verhulst & Akkerhuis, 1986) on a representative sample of 4-5-year-olds from the general population of the same Dutch province of Zuid-Holland (the 1983 sample). Mean scores on the CBCL/4-18 and the TRF were computed for boys and girls (see Table 8.1), and compared to the normative samples. ANCOVA's were performed on the CBCL and TRF cross-informant scales, Internalizing, Externalizing, and total problem scores in a 2 (sample) x 2 (sex) x 2 (age 4 and 5) factorial design with SES as covariate. Sample sizes for the Verhulst data were 318 for the CBCL (151 boys, 167 girls) and 248 for the TRF (116 boys, 132 girls).

Table 8.1
Means and concurrent correlations of CBCL/4-18 and TRF scale scores by sex[1] at age 4-5 yrs, and odds ratios for deviance on TRF scales for children who were deviant versus non-deviant on CBCL/4-18 scales

	CBCL/4-18		TRF			
Scale	Boys (n=204)	Girls (n=193)	Boys (n=176)	Girls (n=173)	*r*	Odds ratio
Withdrawn	2.26	2.22	2.43	2.18	.26 ^{c}	4.6
Somatic complaints	0.76	0.86	0.35	0.49	.23 ^{c}	7.2
Anxious/Depressed	1.93	1.88	3.08	2.33 ^{a**}	.12	--
Social Problems	1.29	1.04	2.42	1.60 ^{a}	.35 ^{c}	11.4
Thought Problems	0.66	0.47 ^{a**}	0.64	0.36	.28 ^{c}	8.4
Attention Problems	3.36	2.50 ^{c}	7.02	4.06 ^{c}	.39 ^{c}	--
Delinquent Behavior	1.76	0.81 ^{c}	1.02	0.57 ^{b}	.19 ^{c}	7.3
Aggressive Behavior	9.29	6.48 ^{c}	6.54	3.11 ^{c}	.37 ^{c}	20.4
Internalizing	4.89	4.89	5.80	4.92	.16 ^{b}	--
Externalizing	11.06	7.29 ^{c}	7.55	3.67 ^{c}	.37 ^{c}	4.8
Total Problems	26.00	21.14 ^{c}	22.65	14.22 ^{c}	.33 ^{c}	3.3

Note. [1] Significance of sex differences in mean scale scores according to ANCOVAs.
Significance levels: ^{a} $p < .05$; ^{b} $p < .01$; ^{c} $p < .001$.
Asterisks indicate those differences most likely to be chance findings.
Only odds ratios of which the confidence interval excluded 1.0 are displayed.

We found significant ($p < .05$) sample effects on six of the CBCL scores and four of the TRF scores. There were seven significant main effects, two 2-way interactions and one 3-way interaction effect. In the 1991 sample, the CBCL Withdrawn, Somatic Complaints, Thought Problems, and Internalizing scores, and the TRF Somatic Complaints, Anxious/Depressed, and Internalizing scores were higher than in the 1983 sample. Further, the 2-way interactions indicated higher CBCL Delinquent Behavior and Externalizing scores for boys in the 1991 sample, and higher scores for girls in the 1983 sample. The 3-way interaction indicated higher TRF Social Problem scores for 5-year-old boys in

the 1991 sample, but lower scores for 5-year-old girls in the 1983 sample. However, all effects were small according to Cohen's (1988) criteria, only one accounting for 2% of variance (TRF Somatic Complaints), and three accounting for less than 1% of variance.

Small but significant sex differences were found. These are indicated in Table 8.1. Boys had higher Attention Problems, Delinquent Behavior, Aggressive Behavior, Externalizing, and total problem scores on the CBCL than girls. The CBCL Anxious/Depressed score was somewhat higher for 5-year-olds than for 4-year-olds, although this was likely to be a chance finding using a .05 protection level (Sakoda et al., 1954). On the TRF, scores for boys were higher than for girls on the Social Problems, Attention Problems, Delinquent Behavior, Aggressive Behavior, Externalizing, and total problem scales. Delinquent Behavior and Anxious/Depressed were somewhat higher for older children, although this latter difference could have occurred by chance alone. The lack of meaningful sex by sample or age by sample interactions (only one accounted for 1% of variance) indicated that sex- and age-differences in problem scores were similar in both samples of 4-5-year-olds.

Further, to provide a standard to which the level of predictive agreement across parents and teachers could be compared, Pearson correlations were computed for the concurrent relations between CBCL/4-18 and TRF scores at Time 2. Also, odds ratios were computed for concurrent parent ratings on the CBCL/4-18 and teacher ratings on the TRF using the borderline cutoff points as before (T-score $\geq$ 67 for syndrome scales; T-score $\geq$ 60 for Internalizing, Externalizing, and Total Problems). As shown in Table 8.1, the concurrent correlations were all significant at the .01 level or smaller, except for Anxious/Depressed ($p < .05$). Correlations were small to medium according to Cohen's (1988) criteria. The correlations for Somatic Complaints, Anxious/Depressed, Delinquent Behavior, and Internalizing were significantly smaller than those for Attention Problems, Aggressive Behavior, and Externalizing, according to Fisher's z-test. Further, the correlations for Anxious/Depressed, Delinquent Behavior, and Internalizing were smaller than for Social Problems and total problems, and for Anxious/Depressed also lower than for Thought Problems. As was the case in the study reported by McConaughy et al. (1992) the correlations for Social Problems, Attention Problems, Aggressive Behavior, Externalizing, and total problems were all above .30. Different from that study, the correlation for Delinquent Behavior found in this study was below .30 (.19 vs. .39). Odds ratios ranged from nonsignificant for Anxious/Depressed, Attention Problems, and Internalizing to 20.4 for Aggressive behavior. All significant concurrent odds ratios were significant at the .01 level, except for Withdrawn ($p < .02$). Compared to concurrent odds ratios for deviant CBCL en TRF ratings provided by McConaughy et al. (1992), the odds ratios for Somatic Complaints, Social Problems, Thought Problems, and Aggressive Behavior were higher, while the remaining were in a similar range or lower.

These results indicated that both the level of mean CBCL and TRF scores and the level of concurrent quantitative and categorical relations between parent and teacher ratings of problem behavior in the present follow-up sample were well in the expected range.

8.4 Stability of parent ratings of problem behavior

The stability of parent-reported problem behavior over the 2-year period was determined by computing Pearson product-moment correlations between problem scores obtained on the CBCL/2-3 at Time 1 and obtained on the CBCL/4-18 at Time 2. Table 8.2 shows the 2-year stability *r*'s for the seven CBCL/2-3 narrow-band scale scores, Internalizing, Externalizing, and total problem scores with the eight CBCL/4-18 cross-informant syndrome scales, Internalizing, Externalizing, and total problems.

Table 8.2
Correlations between CBCL/2-3 scale scores at age 2-3 yrs and CBCL/4-18 scale scores at age 4-5 yrs

	CBCL/4-18 scale										
CBCL/2-3 scale	Wth	Som	Anx	Soc	Tht	Att	Del	Agg	Int	Ext	TP
Oppositional	.31	.16	.33	.26	.23	.35	.31	.48*	.37	.48*	.51
Withdrawn/Depressed	.36*	.01	.30	.39	.29*	.36*	.20	.19	.33*	.21	.35*
Aggressive	.12	.13	.15	.24	.20	.28	.33*	.41*	.17	.43	.39*
Anxious	.37	.09	.27*	.24	.14	.20	.10	.10	.34*	.11	.25*
Overactive	.13	.08	.22	.24	.22	.47	.21	.41*	.21	.39*	.44*
Sleep Problems	.09*	.12*	.16*	.08*	.15*	.13*	.09*	.13*	.17*	.13*	.20
Somatic Problems	.02	.23	.12*	.06	.05	.06	.07	.06	.15*	.07	.12*
Internalizing	.43	.07	.33*	.35*	.23	.31*	.17	.16	.40*	.17	.34*
Externalizing	.27	.17	.32	.30	.26	.44	.35	.54*	.34	.54*	.56
Total Problems	.35	.20	.39	.38	.31	.49	.36	.49	.43	.50	.60

Note. N = 397. Wth = Withdrawn; Som = Somatic Complaints; Anx = Anxious/Depressed; Soc = Social Problems; Tht = Thought Problems; Att = Attention Problems; Del = Delinquent Behavior; Agg = Aggressive Behavior; Int = Internalizing; Ext = Externalizing; TP = Total Problems. The highest predictive correlation for each CBCL/2-3 score has been underlined. Asterisks indicate which predictive correlations were not significantly different ($p < .05$) from the highest for each CBCL/2-3 score.
Significance levels of r: $p < .01$ with $r = .12$; $p < .001$ with $r = .16$.

Ninety-three of the 110 coefficients were significant at the .01 level. Most coefficients were small to medium, according to Cohen's (1988) criteria. Each of the Time 1 scales, except Sleep Problems and Somatic Problems had at least one medium-level correlation with a Time 2 scale. Only the correlations between Time 1 Oppositional and Time 2 total problem scores, and those between Time 1 Externalizing and Time 2 Aggressive, Externalizing, and total problems, and Time 1 total problems and Time 2 Externalizing and total problems could be considered large ($r \geq .50$). The correlations between Time 1 and Time 2 Internalizing, Externalizing, and total problems were .40, .54, and .60 respectively. Both Externalizing and total problem scores showed a significantly higher stability than Internalizing scores ($p < .05$ according to Fisher's z-test). Although 3-year-olds of both sexes tended to have higher externalizing and total problem scores, these differences were not significant.

To indicate with which specific CBCL/4-18 scale each CBCL/2-3 scale showed the

strongest longitudinal relation, the highest correlation of each CBCL/2-3 scale with a CBCL/4-18 score has been underlined in Table 8.2. Asterisks indicate which predictive correlations were not significantly different ($p < .05$) from the highest correlation for each CBCL/2-3 scale, using Fisher's z-test. Most correlations were in predictable directions. All CBCL/2-3 scales denoting internalizing problems had their highest correlations with CBCL/4-18 scales denoting internalizing problems, and all CBCL/2-3 scales denoting externalizing problems had their highest correlations with CBCL/4-18 scales denoting externalizing problems. However, Withdrawn/Depressed had also relatively high correlations with Social Problems, Attention Problems, and Thought Problems. Further, Somatic Problems had also relatively high correlations with Internalizing and Anxious/Depressed beside its correlation with Somatic Complaints, which was the highest. For each CBCL/2-3 scale the cross-time correlation with the CBCL/4-18 total problem score was the highest or one of the highest of all cross-time correlations.

The persistence of parent-reported problems across the 3-year interval was examined by computing the proportion of children with scale scores considered deviant or nondeviant on the CBCL/2-3 at Time 1 and on the CBCL/4-18 at Time 2, using the borderline criterion for scores on both instruments. The bottom of the borderline range was defined by approximately the 85th percentile on the total problem, Internalizing, and Externalizing score (normalized T-score of 60), and by approximately the 95th percentile on the other scale scores (normalized T-score of 67). Proportions were tabulated for children who had deviant or nondeviant CBCL/2-3 scale scores at Time 1 and who had deviant or nondeviant CBCL/4-18 scores at Time 2. Appendix H.3 gives the proportions of children with deviant CBCL/2-3 scale scores at age 2-3 years, who also had deviant CBCL/4-18 scale scores at age 4-5 years, and the proportions of children with deviant CBCL/4-18 scale scores at age 4-5 years, who also had deviant CBCL/2-3 scale scores at age 2-3 years. Relations between Time 1 and Time 2 deviance were tested using chi-squares. Relative risk odds ratios (Fleiss, 1981) were computed for predicting deviance on the CBCL/4-18 at Time 2 from Time 1 CBCL/2-3 scores (see Table 8.3).

Of the children who had deviant CBCL/2-3 Internalizing, Externalizing, or total problem scores at Time 1, the percentages who had a deviant score on the corresponding Time 2 CBCL/4-18 scale were 31%, 40%, and 49%, respectively. When used as a scree-

Table 8.3
Odds ratios for deviance on Time 2 CBCL/4-18 scales for children who were deviant versus non-deviant on Time 1 CBCL/2-3 scales

		CBCL/4-18 scale												
CBCL/2-3 scale	n	Wth (17)	Som (18)	Anx (18)	Soc (22)	Tht (9)	Att (19)	Del (14)	Agg (20)	Int (55)	Ext (57)	TP (55)	I/E (88)	Any (77)
Oppositional	20	4.6	--	6.5	7.1	--	4.0	5.9	20.3	7.4	11.1	19.1	16.9	7.2
Withdrawn/Depressed	15	10.3	--	6.1	10.7	8.2	5.7	--	5.4	8.1	4.3	11.0	7.8	9.4
Aggressive	17	--	5.2	--	4.1	7.1	--	--	7.0	--	3.5	6.3	4.3	5.2
Anxious	16	6.0	--	--	--	--	--	--	--	5.4	--	3.0	4.9	6.3
Overactive	16	--	--	--	--	--	--	4.4	4.9	--	5.1	4.1	4.9	--
Sleep Problems	18	--	--	--	--	6.6	--	--	--	--	--	--	--	--
Somatic Problems	21	--	4.0	--	--	--	--	--	--	--	--	--	--	--
Internalizing	71	13.1	--	3.1	2.8	--	2.9	3.7	3.3	4.0	2.0	3.6	2.5	2.4
Externalizing	62	3.1	--	6.1	2.7	--	4.4	3.2	16.0	5.0	6.4	8.5	6.9	4.0
Total Problems	61	5.5	--	8.0	4.3	7.4	5.6	6.1	12.7	6.4	5.4	12.0	7.2	4.5
Internalizing and/or Externalizing	112	6.7	--	3.4	--	--	3.8	3.6	11.7	4.6	4.1	6.0	4.3	3.0
Any one or more	85	2.7	--	3.1	2.7	--	--	--	6.2	3.0	3.3	5.2	3.9	3.1

Note. N = 397. Wth = Withdrawn; Som = Somatic Complaints; Anx = Anxious/Depressed; Soc = Social Problems; Tht = Thought Problems; Att = Attention Problems; Del = Delinquent Behavior; Agg = Aggressive Behavior; Int = Internalizing; Ext = Externalizing; TP = Total Problems; I/E = Internalizing and/or Externalizing; Any = Any one or more. Numbers of children having deviant CBCL/4-18 scores are given in parentheses.
[a] Borderline cutpoint: T-score ≥ 67 for syndrome scales; T-score ≥ 60 for Internalizing, Externalizing, and Total problems.

ning test, the CBCL/2-3 total problem score showed a predictive sensitivity of 63%, and a predictive specificity of 84%, for an overall correct prediction of 73%. The predictive power of deviant scores on Time 1 narrow-band scales to Time 2 narrow-band scales was relatively low. However, relatively high proportions of the children with a deviant Time 1 score on Oppositional (80%), Withdrawn/Depressed (67%), Aggressive (53%), Anxious (56%), Overactive (56%), Externalizing (57%), or Total Problems (57%) had a deviant Internalizing and/or Externalizing score at Time 2. Similarly, on average 56.6% (range 33% to 80%) of the children who had a deviant CBCL/4-18 score at Time 2 had had a deviant Internalizing and/or Externalizing score at Time 1.

In Table 8.3 only odds ratios are shown that were significantly greater than 1.0 at $p < .05$ as indicated by confidence intervals. The odds ratios ranged from 2.7 for Any one or more at Time 1 on Time 2 Withdrawn and Social problems, to 20.3 for Time 1 Oppositional on Time 2 Aggressive. That is, children who had been deviant on any one or more CBCL/2-3 scales at Time 1 were 2.7 times more likely to be deviant on the CBCL/4-18 Withdrawn or Social Problems scale than were children who had not been deviant on any Time 1 scale. Likewise, children who had been deviant on the Oppositional scale at Time 1 were 20.3 times more likely to be deviant on the CBCL/4-18 Aggressive scale than were children who had not been deviant on the Time 1 Oppositional scale. For each Time 1 scale, the highest odds ratio closely matched one of the highest Pearson correlations reported in Table 8.2, except for the Overactive scale, where the odds ratio with Time 2 Attention Problems was nonsignificant.

8.5 Predictive relations between Time 1 parent ratings and Time 2 teacher ratings

The predictive relations between parent and teacher ratings were tested by using the same analytic procedures as with the comparison of Time 1 versus Time 2 parent ratings. Pearson product moment correlations were computed between CBCL/2-3 scores at Time 1 and TRF scores at Time 2 for 349 children.

Table 8.4 shows these predictive correlations for the seven CBCL/2-3 narrow-band scale scores, Internalizing, Externalizing, and total problem scores with the eight CBCL/4-18 cross-informant syndrome scales, Internalizing, Externalizing, and Total Problems. Of these correlations, 44.5% were significant at the .01 level, and all were small according to Cohen's (1988) criteria. The correlation between Time 1 and Time 2 Internalizing, Externalizing, and total problems was .11, .25, and .24 respectively. Both Externalizing and total problem scores showed a significantly ($p < .05$) higher predictive correlation than Internalizing scores. The highest correlations for each CBCL/2-3 scale with teacher ratings showed a pattern that was quite similar to that with parent ratings, but the following findings were in contrast with those on the parent ratings. The predictive correlations of Time 1 Withdrawn/Depressed with Time 2 TRF Withdrawn and Internalizing were not significant. The Time 1 Anxious scale had no significant relation

Table 8.4
Correlations between CBCL/2-3 scale scores at age 2-3 yrs and TRF scale scores at age 4-5 yrs

	TRF scale										
CBCL/2-3 scale	Wth	Som	Anx	Soc	Tht	Att	Del	Agg	Int	Ext	TP
Oppositional	.05	.03	.11	.10	.09	.14	.16	.18	.10	.19	.17
Withdrawn/Depressed	.12	.04	.09	.17	.17	.20	.11	.17	.12	.17	.15
Aggressive	.01	-.04	.15	.13	.08	.12	.22	.23	.09	.25	.20
Anxious	.13	-.01	.03	.01	.08	.05	.00	-.03	.08	-.03	.03
Overactive	.03	-.03	.13	.20	.11	.24	.18	.21	.09	.22	.24
Sleep Problems	-.01	-.01	.03	.10	.12	.06	.03	.09	.01	.08	.08
Somatic Problems	-.13	.05	.05	-.01	-.03	-.02	-.07	-.04	-.02	-.05	-.04
Internalizing	.15	.01	.06	.08	.14	.13	.05	.05	.11	.06	.12
Externalizing	.04	-.00	.15	.16	.11	.19	.21	.24	.11	.25	.23
Total Problems	.06	.01	.15	.18	.16	.21	.18	.22	.12	.23	.24

Note. N = 349. Wth = Withdrawn; Som = Somatic Complaints; Anx = Anxious/Depressed; Soc = Social Problems; Tht = Thought Problems; Att = Attention Problems; Del = Delinquent Behavior; Agg = Aggressive Behavior; Int = Internalizing; Ext = Externalizing; TP = Total Problems. The highest significant predictive correlation for each CBCL/2-3 score has been underlined.
Significance levels of r: $p < .01$ with $r = .13$; $p < .001$ with $r = .17$.

with the TRF Anxious/Depressed and Internalizing scales. None of the cross-time correlations of Sleep Problems was significant. The only significant predictive correlation for Somatic Problems was with the TRF Withdrawn scale, and this was negative. The Time 1 Internalizing score had no significant correlation with Time 2 Anxious/Depressed, Social Problems, Internalizing, and Total Problems. Finally, the Time 1 Anxious, Sleep Problems, Somatic Problems, and Internalizing scores had no significant correlation with Time 2 Total Problems, although these were also the lowest correlations with Total Problems using Time 2 parent ratings as a criterion.

The proportions of children having deviant CBCL/2-3 scale scores (above the borderline cutpoint) who also had deviant TRF scale scores, and the proportions of children who had deviant TRF scores and also a deviant CBCL/2-3 score are presented in Appendix H.4. Of the children who were deviant on Internalizing, Externalizing, or total problems at Time 1, 19%, 27%, and 25%, respectively, remained deviant on the corresponding TRF scales at Time 2. As was the case with the parent ratings, relatively high proportions of the children with a deviant Time 1 scale scores had a deviant Internalizing and/or Externalizing score at Time 2, and vice versa. On average 37.8% (range 28% to 57%) of the children with a deviant CBCL/2-3 score had a deviant Internalizing and/or Externalizing score at Time 2, and on average 44.4% (range 35% to 71%) of the children who had a deviant TRF score at Time 2 had had a deviant Internalizing and/or Externalizing score at Time 1.

Table 8.5
Odds ratios for deviance on Time 2 TRF scales for children who were deviant versus non-deviant on Time 1 CBCL/2-3 scales

		TRF scale												
CBCL/2-3 scale	n	Wth (19)	Som (12)	Anx (15)	Soc (15)	Tht (15)	Att (17)	Del (15)	Agg (17)	Int (52)	Ext (52)	TP (53)	I/E (92)	Any (73)
Oppositional	20	--	--	--	--	--	--	--	4.0	--	2.6	--	--	--
Withdrawn/Depressed	14	--	5.4	--	--	--	6.3	--	6.3	4.7	--	4.6	4.0	4.1
Aggressive	15	--	--	--	--	--	--	--	--	3.1	--	--	--	--
Anxious	16	--	--	--	--	--	--	--	--	--	--	--	--	--
Overactive	16	--	--	6.2	--	--	5.3	--	--	--	--	--	--	--
Sleep Problems	16	--	--	--	--	6.2	--	--	--	--	--	--	--	--
Somatic Problems	19	--	--	--	--	--	--	--	--	--	--	--	--	--
Internalizing	65	--	--	--	--	--	5.6	--	--	--	--	--	--	1.9
Externalizing	60	--	--	--	--	--	--	--	3.7	--	--	2.2	2.0	--
Total Problems	56	--	--	--	--	--	--	--	--	2.2	--	2.2	--	--
Internalizing and/or Externalizing	104	--	--	--	--	--	6.3	--	2.8	--	--	2.0	--	1.9
Any one or more	80	--	--	3.1	--	--	--	--	--	--	--	--	--	--

Note. N = 349. Wth = Withdrawn; Som = Somatic Complaints; Anx = Anxious/Depressed; Soc = Social Problems; Tht = Thought Problems; Att = Attention Problems; Del = Delinquent Behavior; Agg = Aggressive Behavior; Int = Internalizing; Ext = Externalizing; TP = Total Problems; I/E = Internalizing and/or Externalizing; Any = Any one or more. Numbers of children having deviant CBCL/4-18 scores are given in parentheses.

[a] Borderline cutpoint: T-score ≥ 67 for syndrome scales; T-score ≥ 60 for Internalizing, Externalizing, and Total Problems.

Most of the odds ratios computed on the predictive categorical deviant/nondeviant classifications were nonsignificant. As shown in Table 8.5, the significant ($p < .05$) odds ratios ranged from 2.2 for Externalizing and Total Problems to 6.3 for Withdrawn. No significant odds ratios were found for Anxious and Somatic Problems. Deviance on the CBCL/2-3 Withdrawn/Depressed scale predicted deviance on five of the eleven TRF scales two years later.

8.6 Prediction of Time 2 parent and teacher ratings of problem behavior

Path analyses were performed to assess the relations among the four sets of predictors and the parent and teacher ratings of the eight cross-informant syndromes. The procedure described by Stanger et al. (1992) was followed. Each analysis involved a series of least squares multiple regressions. In each multiple regression, all candidate predictors for a given outcome were entered simultaneously.

The first regression in each analysis was aimed at the identification of direct predictors, i.e., variables that significantly predicted the outcome with the effects of all other predictors partialled out. The next regressions identified indirect predictors, i.e.,variables that predicted the direct predictors. Variables were considered candidate indirect predictors if they were placed prior to the direct predictors in the path. The sequence of the sets of candidate predictor variables in the path was: Time 1 family and child variables, Time 1 parent ratings of problem behavior, stressful life-events, and Time 2 family and child variables. This sequence was based on the following a priori decisions: (1) all variables measured at Time 1 were tested as predictors of those measured at Time 2; (2) Time 1 family and child variables were tested as predictors of Time 1 parent ratings of problem behavior; and (3) parent reports of stressful life-events were added to Time 1 family and child variables and Time 1 parent ratings of problem behavior as predictors of Time 2 family and child variables, because they covered the 12 months period before the Time 2 assessments.

For each direct predictor, indirect predictors were identified through regressions that included all prior variables in the path. For example, if a Time 2 variable was a direct predictor of a Time 2 syndrome, then Time 1 family and child variables, Time 1 parent ratings of problem behaviors, and Time 2 parent reports of stressful events were all tested simultaneously as predictors of that Time 2 variable. If any of these variables predicted the Time 2 variable, they were regarded indirect predictors of the Time 2 syndrome through their association with that variable.

Our purpose was to analyze the contribution of the CBCL/2-3 scales in the prediction of Time 2 parent and teacher ratings of problem behavior. However, because the Time 1 total problem score also showed sizable relations with most Time 2 parent ratings, first, all analyses were executed using the narrow-band CBCL/2-3 scales as predictors, and subsequently using the total problem score as predictor.

We tested whether the pattern of intercorrelations of candidate predictor variables and

the pattern of zero-order predictive correlations regarding Time 2 parents' and teachers' problem ratings were significantly different ($p < .01$) for boys and girls. The intercorrelations between the Time 1 general development score and Time 2 school competence, between difficult temperament and language problems, and between difficult temperament and school competence were higher for boys than for girls. The predictive correlations between school competence and teacher rated withdrawn behavior, and between parenting stress and teacher rated delinquent behavior were higher for girls, whereas the predictive correlation between the Time 1 Withdrawn/Depressed score and teacher-rated total problems was higher for boys. No significant sex differences in predictive correlations regarding Time 2 parent ratings on the CBCL/4-18 were found. Because these differences were few considering the number of 472 comparisons, all multiple regressions were computed on the combined sample of boys and girls. To account for sex and age differences in outcomes, all regressions were computed controlling for sex and Time 1 age.

To reduce the number of variables in the regressions, the following candidate predictor variables were dropped from the analyses, because they had not a single zero-order correlation with the Time 2 syndromes $\geq .25$: Time 1 Sleep problems and Somatic problems, Time 2 SES, marital status, number of sibs, and number of family members having received mental health services.

The intercorrelations between predictors and the predictive correlations are given in Appendices H.5 and H.6 (except those for the CBCL/2-3 scales; see Tables 8.2 and 8.4).

8.6.1 Prediction of parent and teacher ratings

Table 8.6 shows the results of the multiple regression analyses on Time 2 parent and teacher ratings, and the direct predictors for each syndrome with beta weights that were significant at the .01 level. Table 8.7 shows the beta weights for the relation between direct predictors and variables that were placed prior to these predictors in the model. Although the beta weights do not directly reflect the percent of variance accounted for, the predictor variables accounting for the largest proportion of variance have the highest beta weights. The total percentage of variance accounted for by the full models is given by the R^2 adjusted for the number of variables in the equation. Results of the path analyses for the eight syndromes are described below in the order of the relation between the syndromes and the broad-band Internalizing and Externalizing dimensions, described by Achenbach (1991a). To assess differences in predictability among narrow-band and broad-band ratings, the results of the multiple regressions on Internalizing, Externalizing, and total problems are given in addition.

Table 8.6
Multiple regressions and direct predictors of Time 2 CBCL/4-18 and TRF syndromes

CBCL/4-18 Syndrome	Predictors of CBCL/4-18 Syndrome	β	TRF Syndrome	Predictors of TRF Syndrome	β
Withdrawn	Oppositional	.22	**Withdrawn**	School competence	-.40
	Withdrawn/Depressed	.22			
R = .54	Anxious	.19	R = .47		
R^2-adj = .26	Overactive	-.15	R^2-adj = .17		
F = 10.41	Health problems	.11	F = 7.11		
p = .000	Parenting stress	.13	p = .000		
Somatic Complaints	Health problems	.23	**Somatic Complaints**	Health problems	.12
				School competence	-.16
R = .32			R = 26		
R^2-adj = .06			R^2-adj = .03		
F = 2.84			F = 1.82		
p = .001			p = .04		
Anxious/Depressed	General Development	.10	**Anxious/Depressed**	Life-events	.11
	Oppositional	.19		Language problems	-.13
R = .49	Withdrawn/Depressed	.20	R = .43	School competence	-.42
R^2-adj = .21	Life-events	.12	R^2-adj = .15		
F = 8.04	Health problems	.16	F = 5.88		
p = .000	Parenting stress	.12	p = .000		

Table 8.6 (Continued)

CBCL/4-18 Syndrome	Predictors of CBCL/4-18 Syndrome	β	TRF Syndrome	Predictors of TRF Syndrome	β
Social Problems	Withdrawn/Depressed	.32	**Social Problems**	Anxious	-.11
	Health problems	.12		Life-events	.10
R = .61	Language problems	.12	R = .60	School competence	-.50
R^2-adj = .35	School competence	-.18	R^2-adj = .34		
F = 15.11	Parenting stress	.15	F = 14.69		
p = .000			p = .000		
Thought Problems	Disadvantage	-.11	**Thought Problems**	Life-events	.14
	Withdrawn/Depressed	.25		School competence	-.21
R = .43	Difficult temperament	.12	R = .34		
R^2-adj = .16	School competence	-.13	R^2-adj = .08		
F = 5.98			F = 3.35		
p = .000			p = .000		
Attention Problems	Withdrawn/Depressed	.21	**Attention Problems**	Anxious	-.09
	Overactive	.25		Overactive	.11
R = .69	Life-events	.09	R = .70	Life-events	.12
R^2-adj = .45	Difficult temperament	.12	R^2-adj = .46	Language problems	.18
F = 23.15	Language problems	.12	F = 24.07	School competence	-.53
p = .000	School competence	-.13	p = .000		
	Parenting stress	.24			

Table 8.6 (Continued)

CBCL/4-18 Syndrome	Predictors of CBCL/4-18 Syndrome	β	TRF Syndrome	Predictors of TRF Syndrome	β
Delinquent Behavior	Oppositional	.21	**Delinquent Behavior**	Aggressive	.16
	Aggressive	.17		School competence	-.22
R = .46	Life-events	.13	R = .34		
R^2-adj = .18	Language problems	.13	R^2-adj = .08		
F = 6.99	Parenting stress	.11	F = 3.41		
p = .000			p = .000		
Aggressive Behavior	General Development	.11	**Aggressive Behavior**	Aggressive	.15
	Oppositional	.34		Anxious	-.14
R = .66	Aggressive	.13	R = .53	Language problems	-.18
R^2-adj = .41	Anxious	-.13	R^2-adj = .25	School competence	-.45
F = 19.51	Life-events	.10	F = 10.00		
p = .000	Health problems	.10	p = .000		
	Difficult temperament	.17			
	Parenting stress	.14			
Internalizing	Oppositional	.24	**Internalizing**	School competence	-.48
	Withdrawn/Depressed	.19			
R = .56	Anxious	.14	R = .49		
R^2-adj = .29	Life-events	.10	R^2-adj = .21		
F = 11.75	Health problems	.21	F = 8.14		
p = .000	Parenting stress	.13	p = .000		

Table 8.6 (Continued)

CBCL/4-18 Syndrome	Predictors of CBCL/4-18 Syndrome	β	TRF Syndrome	Predictors of TRF Syndrome	β
Externalizing	General Development	.11	**Externalizing**	Aggressive	.16
	Oppositional	.34		Anxious	-.14
R = .65	Aggressive	.15	R = .53	Language problems	-.17
R^2-adj = .40	Anxious	-.13	R^2-adj = .25	School competence	-.44
F = 19.19	Life-events	.11	F = 9.86		
p = .000	Health problems	.09	p = .000		
	Difficult temperament	.15			
	Parenting stress	.15			
Total Problems	General Development	.12	**Total problems**	Anxious	-.11
	Oppositional	.30		Life-events	.10
R = .73	Withdrawn/Depressed	.18	R = .66	School competence	-.58
R^2-adj = .51	Life-events	.14	R^2-adj = .41		
F = 28.82	Health problems	.13	F = 19.54		
p = .000	Difficult temperament	.17	p = .000		
	Parenting stress	.17			

Note. N = 348.

Table 8.7
Indirect predictors* of Time 2 CBCL/4-18 and TRF syndromes

Direct predictor	Indirect predictor	β
Time 1 Oppositional	Disadvantage	.40
Time 1 Withdrawn/Depressed	Disadvantage	.24
	General development	-.18
Time 1 Aggressive	Disadvantage	.33
Time 1 Anxious	Disadvantage	.19
	General development	-.13
Time 1 Overactive	Disadvantage	.30
	General development	-.18
Life-events	--	
Health problems	Disadvantage	.13
Difficult temperament	General development	-.10
	Anxious	.16
	Overactive	.24
	Life-events	.12
Language problems	General development	-.16
School competence	General development	.12
	Withdrawn/Depressed	-.13
Parenting stress	General development	-.13
	Oppositional	.14
	Aggressive	.15
	Overactive	.14

Note. Betas were derived from significant ($p < .05$) multiple regressions of preceding predictor variables on direct predictors.

Teacher ratings of school competence stood out as the strongest direct predictor of all Time 2 TRF syndromes with betas ranging from -.16 for Somatic Complaints to -.53 for Attention Problems. This indicates that higher teacher ratings of children's problem behavior were consistently related to low teacher ratings of the school competence of these children. Because it was a consistent direct predictor of all teacher rated syndromes, school competence will not be described as a direct predictor for each separate teacher rated syndrome. There was no variable that stood out as a direct predictor of all parent rated syndromes.

Withdrawn

There were six direct predictors of Time 2 parent ratings of the Withdrawn syndrome: Time 1 Oppositional, Time 1 Withdrawn/Depressed, Time 1 Anxious, Time 1 Overactive, health problems, and parenting stress. The beta weights of .22 for both the Time 1 parent ratings on the Oppositional and the Withdrawn/Depressed scale indicate that these were equally strong predictors of the Time 2 Withdrawn syndrome. The negative beta weight of -.15 for Time 1 Overactive indicated that young preschoolers showing much overactive behavior are *less* likely to display withdrawn behavior in preschool than other children.

School competence was the only direct predictor of teacher ratings on the Time 2 Withdrawn syndrome.

Somatic Complaints

Time 2 health problems was the only direct predictor of Time 2 parent ratings of Somatic Complaints, having a beta of .23.

In addition to school competence, the Time 2 teacher ratings of Somatic Complaints had Time 2 health problems as a direct predictor.

Anxious/Depressed

Time 1 general development, Oppositional, and Withdrawn/Depressed, life-events between Time 1 and Time 2, and Time 2 health problems and parenting stress were direct predictors of parents' Time 2 ratings of the Anxious/Depressed syndrome. Of these direct predictors Time 1, Oppositional and Withdrawn/Depressed had the highest betas (.19 and .20, respectively).

Time 2 Anxious/Depressed scores based on teacher ratings were directly associated with life-events reported by parents. Language problems was also a direct, but negative predictor of teachers' Anxious/Depressed ratings, indicating that higher anxious/depressed scores tended to go together with *fewer* language problems.

Social Problems

Time 1 parent ratings of Withdrawn/Depressed was the strongest predictor of Time 2 parent ratings of Social Problems, having a beta of .32. Beside these, Time 2 parent reports of health problems, language problems, and parenting stress, and teacher ratings of school competence were included as direct predictors of Time 2 Social Problems as rated by parents.

Direct predictors of teacher ratings of Social Problems were life-events and Time 1 parent ratings of anxious behavior. The beta of anxious was negative, indicating that children with high Anxious scores at Time 1 were *less* likely to show social problems in the classroom.

Thought Problems

Parent ratings of Thought Problems had Time 1 Withdrawn/Depressed ratings as the strongest direct predictor variable (beta=.21). Additional direct predictors were Time 1 disadvantage, and Time 2 difficult temperament and school competence. The negative beta for disadvantage indicated that higher Time 2 Thought Problems scores were related to *lower* disadvantage scores at Time 1.

Life-events reported by the parent at Time 2 were the only direct predictor of teacher rated thought problems beside the school competence score.

Attention Problems

Time 2 parent ratings of Attention Problems had seven direct predictors, of which Time 1 Overactive (beta=.25), Time 2 parenting stress (beta=.24), and Time 1 Withdrawn/Depressed (beta=.21) were the strongest. The other four direct predictors were life-events, difficult temperament, language problems, and teacher rated school competence.

Beside the teacher ratings of school competence, also Time 1 parent ratings of anxious and overactive behavior, and life-events and Time 2 language problems contributed directly to the prediction of Time 2 teacher rated Attention Problems. The negative sign of the beta for Time 1 Anxious showed that highly anxious toddlers were *less* likely to show attention problems at school two years later.

Delinquent Behavior

Time 2 parent ratings of Delinquent Behavior were directly predicted by Time 1 parent ratings of oppositional and aggressive behavior, which were the strongest predictors with betas of .21 and .17, and by Time 2 reports of life-events, language problems, and parenting stress.

Apart from school competence was Time 1 Aggressive the only direct predictor of teacher ratings of delinquent behavior.

Aggressive Behavior

Parents' Time 2 ratings of the Aggressive Behavior syndrome were directly associated with most of the predictor variables. These were Time 1 parent ratings of general development, oppositional, aggressive, and anxious behavior, Time 2 reports of life-events, health problems, difficult temperament, and parenting stress. Of these, Time 1 oppositional behavior was the strongest direct predictor (beta=.34). Higher Aggressive Behavior scores at Time 2 were associated with *higher* scores on Time 1 general development and with *lower* scores on the Time 1 Anxious scale.

Apart from school competence, there were three direct predictors of Time 2 teacher ratings of Aggressive Behavior: the Time 1 Aggressive and Anxious scales, and Time 2 language problems. Higher scores on the teacher rated Aggressive Behavior syndrome were associated with *lower* Time 1 Anxious scores, and with *less* language problems at Time 2.

8.6.2 Indirect predictors

As shown in Table 8.7, all predictors included prior to direct predictors in the regression models of Time 2 syndromes were associated with one or more direct predictors, making them indirect predictors of Time 2 parent or teacher ratings. Both Time 1 family and child variables were associated with Time 1 and Time 2 predictors. Time 1 disadvantage was related to each of the Time 1 problem scales with betas ranging from .19 to .40, and also with Time 2 health problems (beta=.13). Time 1 general development was negatively related to Time 1 Withdrawn/Depressed, Time 1 Anxious, and Time 1 Overactive, and to difficult temperament, language problems, and parenting stress at Time 2, with betas ranging from -.10 to -.18. The Time 1 general development was positively related to Time 2 school competence (beta=.12). Besides being direct predictors of Time 2 parent problem ratings, Time 1 Oppositional and Aggressive predicted parenting stress at Time 2 (betas were .14 and .15); Time 1 Withdrawn/Depressed was negatively related to school competence (beta=-.13); Time 1 Anxious was predictive of Time 2 difficult temperament (beta=.16); and Time 1 Overactive predicted difficult temperament (beta=.24) and parenting stress (beta=.14). The life-events score was an indirect predictor through its association with difficult temperament (beta=.12). It was not related itself to any specific Time 1 scale score or Time 1 family or child variable.

8.6.3 Comparison of complete models

As shown in Table 8.6 the percent of variance accounted for by the complete regression models was larger in the prediction of parent ratings of syndromes than in the prediction of teacher ratings. The percent of variance accounted for in parent ratings ranged from 6% for Somatic Complaints to 45% for Attention Problems. All effect sizes except for Somatic Complaints were medium to large according to Cohen's (1988) criteria. Effect sizes for teacher ratings were small for Somatic Complaints (3%), Thought problems (8%), and Delinquent Behavior (8%), and medium to large for the other syndromes, ranging from 15% to 46%.

All multiple regressions were replicated using the Time 1 total problem score as predictor instead of the narrow-band scales. The percentage of variance in each Time 2 parent or teacher rated syndrome, accounted for by the models in which Time 1 total problems replaced the Time 1 narrow-band scales was equal to or smaller than the percentage, accounted for by the models reported in Table 8.6, except for the teacher rated Withdrawn syndrome (17% vs. 19%).

Comparison of the results of the multiple regressions on the Internalizing, Externalizing, and Total Problems score shows that large percentages (cf. Cohen, 1988) of variance in each score were accounted for by the models including the Time 1 narrow-band scales (see Table 8.6). This percentage was low for Internalizing (29%) compared to those for Total problems (51%) and Externalizing (40%). In each of the models Time 1

Oppositional had the highest beta. Comparison of the direct predictors of both broad-band scales shows that Time 1 Oppositional, life-events, health problems, and parenting stress were direct predictors of both. Time 1 general development, Aggressive, and Time 2 difficult temperament were unique predictors of Externalizing, whereas Time 1 Withdrawn/Depressed was a unique predictor of Internalizing. Also Time 1 Anxious was a direct predictor of both broad-band scales. However, while the association with Internalizing was positive, the relation with Externalizing was negative.

To assess the value of Time 1 Total Problems in predicting Time 2 Total problems, a separate path analysis was done using the same predictors and procedure as reported above, but with replacement of the Time 1 narrow-band scales by Time 1 Total Problems. Figure 8.1. shows the relations among the four sets of predictor variables and the Time 2 parent ratings of total problems.

There were seven direct predictors of Time 2 Total Problems, of which parent ratings of Time 1 Total Problems had the highest beta (.46). In addition to Time 1 Total Problems, parents' reports of Time 1 general development, life-events between Time 1 and Time 2, and Time 2 health problems, difficult temperament, and parenting stress, and teacher ratings of school competence were also direct predictors of Time 2 Total Problems. The beta weight for Time 1 Total Problems indicated that it was a stronger predictor of Time 2 Total Problems than any of the other direct predictors for which betas ranged between -.09 and .19. The Time 1 direct predictors Total Problems and general development were also influential as indirect predictors through their association with other direct predictors. Time 1 Total Problems was associated with parents' reports of life-events, difficult temperament, and parenting stress, and with teacher rated school competence at Time 2. Time 1 general development was related with parent rated difficult temperament and teacher rated school competence. In addition, Time 1 disadvantage worked as an indirect predictor of Time 2 Total Problems through its association with Time 1 Total Problems and Time 2 health problems.

8.7 Prediction of signs of disturbance at home and school

There were 348 children with complete data on the parent and teacher reported signs of significant problems. For 136 (41.7%) of these 326 children one or more sign of significant problems was reported. One sign was reported for 70 (51.5%), two signs for 34 (25%), three signs for 17 (12.5%), and four or more signs for 15 (11%) of the children with one or more sign of significant problems.

Of the 136 children for whom one or more signs of disturbance was reported, 76

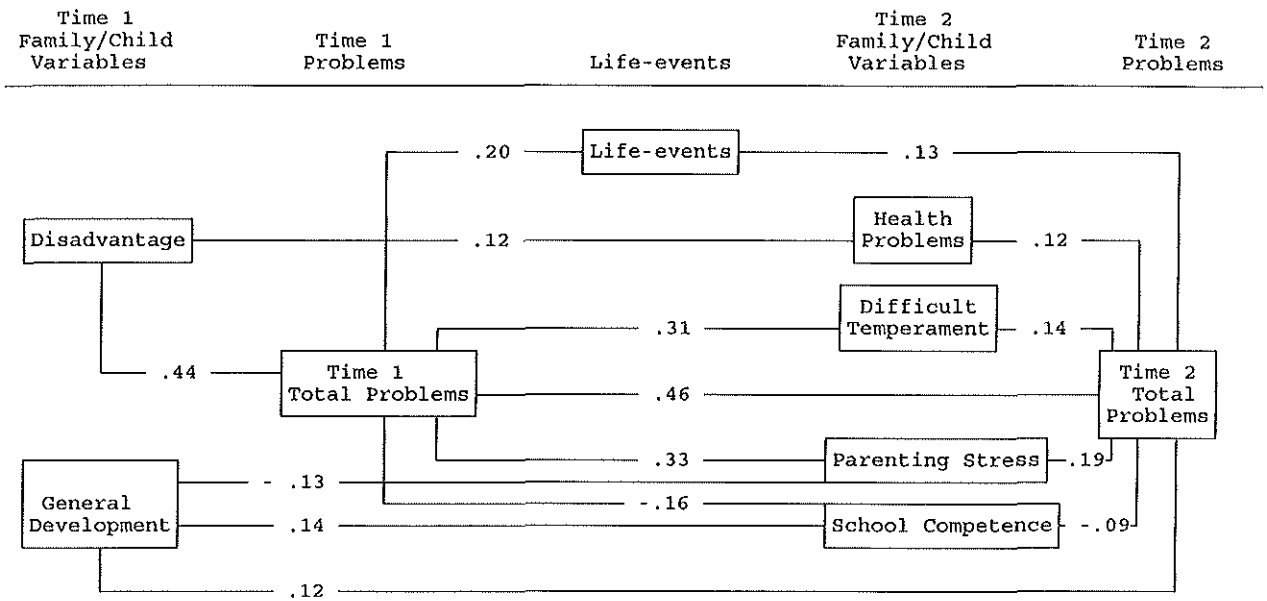


Figure 8.1 Predictors of Time 2 Total Problems

were judged by the parent to have a significant behavioral/emotional problem; 78 were judged by the teacher to have a significant behavioral/emotional problem; 56 were reported by the parent to have received mental health services or special help for academic problems in school; 39 were reported by the teacher to have received mental health services or special help for academic problems in school; 26 were judged by the parent to need professional help for behavioral/emotional problems that was not being received; and 16 were judged by the teacher to need professional help for behavioral/emotional problems that was not being received. Of the children who had received mental health services and special help according to parents, 17.9% had received these from mental health professionals, 64.3% from school related services, and 17.9% from others. These proportions were 15.4%, 59.0%, and 25.6%, respectively, for children reported by teachers to receive mental health services or special help.

To obtain a total disturbance score, the signs of significant problems were summed for each individual child. The same approach to path analysis used for the assessment of the relation between predictors and Time 2 syndromes was chosen to assess the relations among the four sets of predictors and the total disturbance score. Because chi-square tests showed that the distribution of children for whom one or more sign of disturbance were reported was significantly different ($p < .05$) across sex (58.4% boys vs. 41.6% girls) and age (61.3% age 3 vs. 38.7% age 2), sex and age were entered in the model first.

There were five direct predictors of the Time 2 total disturbance score: Time 1 Withdrawn/Depressed (beta=.11), Time 1 Anxious (beta=-.11), parent reports of life-events (beta=.15), teacher reported school competence (beta=-.19), and language problems (beta=.25). Higher Time 1 Anxious scores were apparently related to *lower* total disturbance scores at Time 2. The full model accounted for 31% of the variance in the Total Disturbance score.

8.8 Differences between children with persisting, increasing, and decreasing problems

The multiple regression analyses reported above answered the question to what extent the rank ordering of children's problem scores at ages 4-5 obtained from parents and teachers may be predicted from problem scores obtained from parent ratings two years earlier. They do not answer questions regarding change. To explore which child and family characteristics were associated with change, we analyzed differences on predictor variables among children who's problem scores were persistent, increasing, or decreasing across the follow-up period. From the sample of 348 children for whom complete data were available at both times of assessment, three subsamples were selected using the borderline cutoff points on the Time 1 CBCL/2-3 total problem score and the Time 2 CBCL/4-18 total problem score, i.e., scores above approximately the 85th percentile of the frequency distribution of total problem scores for each sex. One subsample (n = 47) consisted of children who were scored above the borderline cutpoint at both Time 1 and Time 2 and are called "*persisters*". The second subsample (n = 41) consisted of children called "*increasers*" who had scores below the cutpoint at Time 1 and above the cutpoint

at Time 2. The third subsample (n = 65) consisted of children with scores above the cutpoint at Time 1 and below the cutpoint at Time 2. These are called "*decreasers*". The distribution of these three groups across sex, Time 1 age, and Time 1 SES is given in Table 8.8. The distribution of the groups across sex and age did not differ significantly, although persisters seemed somewhat more likely to be 3 than 2 years of age, and increasers seemed to be more likely male than female. However, 55.3% of the persisters came from the lower socio-economic level as compared to 22.0% of the increasers and 30.8% of the decreasers, which is a significant difference (χ^2=13.31, p<.01). ANCOVAs were computed on the predictor variables (see Table 8.9) and Time 2 parent ratings for the three groups with SES as covariate and post-hoc Student-Newman-Keuls (p<.05) comparisons. Although it was considered desirable to include a comparison group drawn randomly from the remaining children in the analyses, preliminary analyses showed that the variances of most predictor variables were not homogeneous for the children in the subsamples and remaining sample.

Table 8.8
Distribution of children with persisting, increasing, and decreasing problems across sex, age, and SES

		Persisters (n = 47)	Increasers (n = 41)	Decreasers (n = 65)
Sex				
	Boys	55.3	61.0	50.8
	Girls	44.7	39.0	49.2
Age				
	Age 2	40.4	56.1	43.1
	Age 3	59.6	43.9	56.9
SES				
	Low	55.3	22.0	30.8
	Middle	23.4	39.0	43.1
	High	21.3	39.0	26.2

Note. Entries indicate proportions of children in each group.

Significant differences between the subsamples were found for eight of the 14 predictor variables. Persisters had higher Time 1 Oppositional, Withdrawn/Depressed, Anxious, Overactive, and total problem scores than increasers, but also higher Time 1 Oppositional and total problem scores than decreasers. The mean Time 1 total problem score for persisters of 57.0 was very close to the mean of 58.1 found for 2-3-year-olds referred to community mental health and child psychiatric services (see Chapter 7), whereas the mean for increasers of 33.2 was identical to the mean of 33.4 found in the

Table 8.9
Means of predictor variables for children with persisting, increasing, and decreasing problems

Predictors	Persisters (n = 47) Mean (SD)	Increasers (n = 41) Mean (SD)	Decreasers (n = 65) Mean (SD)	F (3,150)	p	Contrast
Time 1						
Disadvantage	2.15 (0.82)	1.46 (1.00)	1.78 (0.94)	2.73	.069	--
General development	52.81 (26.32)	58.90 (30.65)	46.46 (31.53)	2.08	.129	--
Oppositional	18.81 (4.90)	11.22 (4.34)	14.08 (5.06)	26.70	.000	P > D > I
Withdrawn/Depressed	3.13 (3.42)	0.71 (1.08)	2.27 (1.96)	11.65	.000	P,D > I
Aggressive	5.13 (3.17)	3.76 (2.33)	4.65 (3.21)	1.75	.177	--
Anxious	5.87 (3.46)	2.68 (2.23)	5.86 (3.29)	15.86	.000	P,D > I
Overactive	5.36 (2.53)	3.07 (2.39)	4.45 (2.69)	6.86	.001	P,D > I
Total Problems	56.98 (13.12)	33.15 (12.82)	46.34 (12.48)	34.29	.000	P > D > I
Time 2						
Life-events	2.05 (2.10)	1.21 (1.38)	1.22 (1.34)	3.38	.037	P > D
Health problems	2.13 (0.87)	2.17 (0.77)	1.71 (0.73)	5.85	.004	P,I > D
Temperament	2.66 (1.66)	1.85 (1.57)	1.74 (1.37)	4.96	.008	P > I,D
Language problems	13.22 (2.31)	11.87 (1.63)	12.96 (2.32)	2.75	.068	--
School competence	-1.22 (2.61)	-0.21 (2.88)	-0.50 (3.28)	0.47	.623	--
Parenting stress	27.23 (8.94)	26.05 (10.93)	22.44 (9.26)	3.03	.052	--

Note: Differences were tested using ANCOVAs with SES as covariate.
Post-hoc contrasts were tested using the Scheffé procedure.

community sample (see Chapter 6). The position of the mean total problem score for decreasers was virtually between these two.

Further, persisters experienced significantly more stressful life-events than decreasers between Time 1 and Time 2; both persisters and increasers had more health problems than decreasers; and persisters were rated as having a more difficult temperament than both increasers and decreasers.

By definition, the mean Time 2 total problem score for persisters of (M=43.7) was higher than for decreasers (M = 22.5), but it was also higher than the mean for increasers (M = 37.7). Persisters had significantly higher scores on all Time 2 CBCL/4-18 syndromes than decreasers, except on Thought Problems. Increasers had significantly higher scores than decreasers on all CBCL/4-18 syndromes, except Thought Problems and Attention Problems. However, increasers had *lower* scores on Time 2 Attention Problems and Aggressive Behavior scores than persisters.

To assess to what extent the deviant behavior of the children in the three subsamples was perceived by parents and teachers to constitute a significant problem, proportions of children in each group having a significant problem were compared. According to parents, 64.5% of the persisters, 41.5% of the increasers, and 29.2% of the decreasers had a significant problem, which is a significant difference between groups (χ^2=13.38, p=.001). Teachers' ratings of significant problems were not significantly different for the three groups (39.0%, 35.3%, and 28.1%, respectively). The proportion of children having a significant problem according to the parent, the teacher, or both was 75.6% for persisters, 55.9% for increasers, and 42.1% for decreasers, which is a significant difference (χ^2=10.89, p=.004).

8.9 Discussion

In this chapter, findings were reported from a 2-year reassessment of the 1989 community sample of 2-3-year-olds, testing the stability of parents' ratings of children's problems; the persistence of deviance reported by parents; the predictive relations between parent ratings at Time 1 and teacher ratings at Time 2; the predictive relations between Time 1 parent ratings and Time 2 parent and teacher ratings, and signs of significant problems independent of Time 1 and Time 2 family and child characteristics, and parents' reports of intervening stressful events; and exploring differences between children with persisting and changing problems.

This longitudinal study was not aimed at reaching conclusions about the etiology of behavioral/emotional problems in preschoolers. Rather, it aimed at the assessment of predictive relations of indications of problem behavior at 2 and 3 years of age with behavioral/emotional problems at school entry, i.e., at age 4-5.

Stabilities of parent ratings

Time 1 parent ratings of CBCL/2-3 scales significantly predicted Time 2 parent ratings of CBCL/4-18 scales. All Time 1 scales except Sleep Problems and Somatic Problems

accounted for medium amounts of variance in Time 2 scale scores. Cross-time correlations of Externalizing and total problem scores were large, and showed significantly larger stabilities for these scores than for Internalizing scores.

These findings were supported by categorical analyses. Of the children who had deviant CBCL/2-3 Internalizing, Externalizing, or total problem scores at Time 1, the percentages who had a deviant score on the corresponding Time 2 CBCL/4-18 scale were 31%, 40%, and 49%, respectively. When used as a screening test, the CBCL/2-3 total problem score showed a predictive sensitivity of 63%, and a predictive specificity of 84%, for an overall correct prediction of 73%. Fifty-three to 80 percent of the children with a deviant Time 1 score on each CBCL/2-3 scale, except Sleep Problems and Somatic Problems, had a deviant CBCL/4-18 Internalizing and/or Externalizing score. On average almost 57 percent of the children who had a deviant CBCL/4-18 score at Time 2 had had a deviant Internalizing and/or Externalizing score at Time 1. Odds ratios showed several strong relations between Time 1 deviant scores on the Oppositional, Withdrawn/Depressed, Internalizing, Externalizing, and total problem scales and Time 2 deviant scale scores on the Withdrawn, Social Problems, Aggressive, Externalizing, and total problem scales.

The overall predictive power of broad-band Internalizing and Externalizing scales, and the total problem scale found in this study was somewhat lower than reported for older children across comparable follow-up periods (McConaughy et al., 1992; Verhulst & Althaus, 1988), but well within the range of continuities reported for population-based follow-up studies of preschoolers (Cohen & Bromet, 1992; Kohn & Rosman, 1972; Richman et al., 1982). The higher stability for Externalizing than for Internalizing problems has also been found in other follow-up studies on preschool samples (Fischer at al., 1984; Kohn & Rosman, 1972). Using the CBCL/2-3 and CBCL/4-18, Achenbach (1992) reported somewhat higher stabilities for the 1992 American version of the CBCL/2-3 scales, whereas Rose et al. (1989) reported lower stabilities for internalizing and total problems than found in the present study. However, these studies are not entirely comparable to ours, because in the Achenbach study almost one-third of the sample were low-birthweight children, while the Rose et al. study included fewer than 40 subjects. Similar to Achenbach (1992) we found consistent, but not significantly lower stabilities for 2-year-olds than for 3-year-olds. Like most other studies we found no differential stability for boys and girls.

Predictive relations between parents and teachers

Predictive correlations between Time 1 parent ratings and Time 2 teacher ratings were low, if significant. However, they showed an overall pattern similar to that found for stabilities of parent ratings. Predictive relations were lower for Internalizing than for Externalizing behavior. Categorical analyses showed that deviant Time 1 parent ratings predicted deviant Time 2 teacher ratings in only 21%-27% of the cases using Internalizing, Externalizing, or total problem scores. However, as was the case with the parent ratings, relatively high proportions of the children with deviant Time 1 scale scores had a

deviant Internalizing and/or Externalizing score at Time 2 (on average 38%), and vice versa (on average 44%). None of the odds ratios between Time 1 deviant parent ratings and Time 2 deviant teacher ratings reached 10.0. Of all CBCL/2-3 scales the Withdrawn/Depressed scale was most predictive of deviance on TRF scales across the two-year period.

The moderate cross-time stability of Time 1 versus Time 2 parent ratings and the low predictive relation between parent and teacher ratings found in this study may have been caused by at least two factors apart from the apparent instability or change in young children's behavior. First, there were differences in the item content of Time 1 and Time 2 checklists. Although the CBCL/2-3 has been designed to tap similar behaviors as the CBCL/4-18 and the TRF, only half of its items have counterparts in the CBCL for older children, and even less recur in the TRF. Given this large difference in item content, the correlation of .60 between Time 1 and Time 2 parent ratings should be considered very substantial. Second, meta-analyses (Achenbach et al., 1987) have shown higher concurrent agreement between similar informants than between different types of informants. Table 8.1 showed concurrent cross-informant correlations that were well in the expected range. Given the two-year time interval it would be surprising to find predictive cross-informant correlations above .30. Even using highly comparable instruments at both times of measurement, McConaughy et al. (1992) found predictive correlations higher than .30 across a 3-year period between parent and teacher ratings for only four of the eight cross-informant syndromes (Achenbach, 1991b).

Longitudinal correlations of specific syndromes

The specificity of longitudinal correlations between Time 1 and Time 2 narrow-band scales was not particularly impressive compared to those found for older children (McConaughy et al., 1992; Verhulst & Althaus, 1988; Verhulst et al., 1990). Although all Time 1 parent rating scales had significant predictive correlations with Time 2 parent ratings in expected directions, at the level of narrow-band syndromes only the Time 1 Oppositional score was differentially predictive of the Time 2 Aggressive Behavior scale at both the quantitative and the categorical level. The correlational patterns for the other scales, except Sleep Problems and Somatic Problems, indicated that scales reflecting internalizing behavior were predictive of later internalizing narrow-band and broad-band scales, and scales reflecting externalizing behavior were predictive of later externalizing narrow-band and broad-band scales. Of the CBCL/2-3 narrow-band scales Oppositional and Withdrawn/Depressed were most predictive of problem behavior two years later. These findings were corroborated by the patterns of association found in the categorical analyses of CBCL scores, and clearly, but much less strongly by the findings on the TRF. We may conclude that, although several narrow dimensions can reliably and validly be distinguished in 2-3-year-old children's problem behavior as reported by parents (see Chapters 4 and 5), these dimensions are of limited value in predicting specific narrow-band syndromes two years later. However, the narrow-band dimensions distinguished at

early preschool age seem more predictive of specific broad-band dimensions at Kindergarten age. Children with stable behavior problems of any kind were to a considerable degree "caught" by having deviant scores on the broad-band Internalizing or Externalizing dimension (or both) using either parent or teacher ratings. Predictive validity defined in this way was high compared to other longitudinal studies on both preschoolers and older children.

Prediction of parent and teacher ratings

Path analyses showed that all CBCL/2-3 scales included in the analyses (excluding Sleep Problems and Somatic Problems) significantly predicted Time 2 CBCL syndromes over and above Time 1 and Time 2 family and child variables, and stressful experiences preceding the Time 2 measurement. For each syndrome, except Somatic Complaints, one or more CBCL/2-3 scales were the strongest of all direct predictors resulting from the regression analyses. Of these, the Oppositional and Withdrawn/Depressed scale functioned as very broad predictors. Both were the strongest predictors of Time 2 Withdrawn and Anxious/Depressed. Further, Oppositional was the strongest predictor of Delinquent Behavior and Aggressive behavior, while Withdrawn/Depressed was the strongest predictor of the Social Problems and Thought Problems syndrome. The Overactive scale was the strongest predictor of Attention Problems, which also had the Time 1 Withdrawn/Depressed scale as the second strongest predictor. Thus, strong oppositional behaviors at age 2-3 tended to be predictive of both internalizing and externalizing problems two years later. Withdrawn/depressed behaviors at this young age tended to predict internalizing problems as well as problems covered by the syndromes that are not strong representatives of either broad-band grouping. Although the CBCL/2-3 Aggressive and Anxious scales showed predictive relations in expectable directions, they were in no case the strongest direct predictor of a Time 2 syndrome.

Time 1 family and child variables were only weak predictors of a few Time 2 syndromes. High Time 1 disadvantage scores predicted lower Thought Problems scores, while a higher general development tended to be predictive of both more anxious/depressed and aggressive behaviors at Time 2.

A higher number of stressful life-events reported for the period of 12 months preceding Time 2 was a weak direct predictor of the Anxious/Depressed, Attention Problems, Delinquent Behavior, and Aggressive Behavior syndromes. The life-events score was predictive of Time 2 total problems even with the influence of Time 1 total problems partialled out, accounting for approximately 3% of variance in the Time 2 total problem score. This is in congruence with findings reported by Berden (1992), but in contrast to the finding of Stanger et al. (1992) that parents' reports of subjective stressful experiences were, but parent reports of "objective" stressful life-events were not predictive of CBCL syndromes across a three-year period. This may be explained by the shorter period prior to Time 2 on which parents were asked to report, employed in this study, which possibly increased the reliability of their reports.

As was expected (see Chapter 2), Time 2 parent reports of health problems, difficult

temperament and parenting stress, and language problems and teacher rated school competence were all related to Time 2 parent ratings of behavioral/emotional problems. However, despite medium to high concurrent correlations with the Time 2 syndromes, in only two instances these predictors were among the strongest for a given syndrome. Time 2 health problems was the only significant direct predictor of Time 2 Somatic Complaints. Parenting stress was the strongest predictor of Attention Problems, together with Time 1 Overactive and Withdrawn/Depressed. The important point here is that for seven out of eight Time 2 parent ratings of syndromes, the Time 1 CBCL/2-3 scales were the strongest predictor over and above family characteristics and other child characteristics, even if these were measured concurrent to the Time 2 syndromes.

The strongest correlates of teacher ratings of children's problem behavior were teacher ratings of school competence as measured by task-related behavior, social behavior, affect, and self-help skills. Apparently, a large part of the variance in preschool teachers' ratings of pupils' behavioral/emotional problems is accounted for by the perceived level of child competence regarding school tasks. A similar result was reported by Kohn and Rosman (1972) who obtained an average correlation of .77 between social competence factors and symptom factors of the Kohn Symptom Checklist. However, over and above teacher rated school competence, three CBCL/2-3 scales had an independent contribution to some TRF syndromes. Time 1 Overactive was a direct predictor of teacher rated Attention Problems, while Time 1 Aggressive directly predicted the TRF Delinquent Behavior and Aggressive Behavior syndromes.

A remarkable finding was that parent reported "objective" life-events also had an independent contribution to teacher ratings on the Anxious/Depressed, Social Problems, Thought Problems, and Attention Problems syndromes, and of the TRF total problem score. This corroborates the suggestive evidence from the analyses of parent ratings that stressful life-experiences have a direct contribution to the development of behavioral/emotional problems, especially of an anxious/depressed and attentional nature.

Prediction of reports of significant problems

Unexpectedly, Time 1 parent ratings of problem behavior hardly contributed to the prediction of Time 2 parents' and teachers' reports of significant problems. Of the Time 1 parent ratings only Withdrawn/Depressed had a positive, and Anxious had a negative contribution. More substantial contributions came from parents' reports of life-events, teacher reported school competence, and language problems. The larger proportion of variance in total disturbance accounted for by the latter two variables may reflect the large proportion of school related problems that were reported by both teachers and parents as signs of disturbance.

Persisters, increasers, and decreasers

The results suggested continuity of behavior problems for some but not all preschool children. In about half of the children considered deviant at one time problem behavior

may be temporary and attributable to concurrent adverse circumstances or a period of developmental transition or adaptation. Comparison of the subsample of children who had persistently high problem scores across the two-year interval ("persisters") with children who's initially high scores decreased considerably ("decreasers"), and with children who's initial scores increased considerably ("increasers") showed that persisters came significantly more often from the lower socioeconomic status group than increasers and decreasers. Independent of SES, persisters had higher mean initial total problem and Oppositional scores, and higher mean life-events, health problems, and difficult temperament scores at Time 2 than decreasers. Compared to increasers, they had higher Time 2 scores on difficult temperament, and also had higher total problem scores, and more specifically, showed more attention problems and aggressive behavior at Time 2. Not surprisingly, three out of four persisters were considered by their parents and/or teachers to have a significant problem. The only difference between increasers and decreasers that was not related to Time 1 or Time 2 problem scores was that increasers showed more health related problems at Time 2.

It seems that those 2-3-year-olds who initially show problems at a high level, comparable to that of children referred for mental health and child psychiatric services, and who are more likely to come from lower SES groups, have a difficult temperament, poor health, and to experience more stressful life-events may be especially at risk for having behavioral/emotional problems at age 4-5. Apart from other problems, these children are likely to show more attention problems and aggressive behavior than 4-5-year-old age-mates who also have high levels of problem behavior.

In summary, we may conclude that (a) the stability of parents' ratings of young preschoolers' problems is for most problem scales medium to high, and lower for internalizing than for externalizing problems; (b) the persistence of behavioral deviance reported by parents is comparable to but less specific than that found for older children; (c) the predictive relations between parent ratings at Time 1 and teacher ratings at Time 2 are low, but reflecting the same pattern as the predictive relations between parent ratings; (d) parent ratings of problem behavior in 2-3-year-olds are predictive of parent and teacher problem ratings two years later independent of family and child characteristics, and parents' reports of intervening stressful events, whereas signs of significant problems independent of high problem ratings are hardly predicted by initial parent ratings; (e) 2-year-olds who have CBCL/2-3 scores comparable to the mean scores of agemates referred to mental health services may be at risk for persistent problems.

The results reported above should be interpreted within the following methodological limitations of the study. First, the longitudinal findings reported in this chapter should not be interpreted as reflecting causal relationships. Findings obtained from the path analyses were based on correlations between predictors and outcomes. No experimental design or apriori contrasts were employed or tested. Findings on differences between changers and persisters necessarily implied post-hoc comparisons.

Further, part of the predictive relations found in this study may reflect stability in parental rating behavior in addition to stability of child behavior and interaction (see Fergusson & Horwood, 1987). However this could not be tested because only at Time 2 information was obtained from multiple informants. However, at Time 2 several cross-informant relations were found. E.g., Time 1 parent ratings, life-events, and Time 2 parenting stress predicted teacher ratings of behavioral/emotional problems, while teacher rated competence accounted for variance in Time 2 parent ratings.

Finally, part of the relation between Time 2 predictors and Time 2 problem ratings may be attributable to early child behavior, either because early child behavior was reflective of the same trait as may be the case with Time 2 difficult temperament, or because it was directly influential on parental feelings, as may be the case with Time 2 parenting stress (see also Anastopoulos, Guevremont, Shelton, and DuPaul, 1992). Because variables such as temperament and parenting stress were not measured at Time 1, the contribution of early temperament and parenting stress to later scores on these variables could not be assessed.

CHAPTER 9

General discussion and conclusions

The main aims of this study were to assess the validity of a parent rating scale for problem behavior in children ages 2-3 years, to obtain data on the prevalence and correlates of problem behavior in young preschoolers from the general population and in a sample referred to mental health services, and to study the stability of problem behavior in preschoolers across a two-year period. In this chapter the main conclusions from the study and some implications for clinical practice and research will be discussed.

9.1 Psychometric characteristics of the CBCL/2-3

The factor structure of the Child Behavior Checklist for Ages 2-3 (CBCL/2-3; Achenbach, Edelbrock, & Howell, 1987; Achenbach, 1992) was investigated with three different samples - children referred to mental health services, children from the general population, and a sample of twin pairs. A series of exploratory and confirmatory factor analyses indicated a seven-factor model for all three samples. The syndrome scales derived from these factors were labeled Oppositional, Withdrawn/Depressed, Aggressive, Anxious, Overactive, Sleep Problems, and Somatic Problems. Factor intercorrelations and a second-order factor analysis provided support for two broad-band groupings of problem behaviors - Externalizing and Internalizing.

Whether the broad-band groups of internalizing and externalizing problems are better regarded as second-order factors or more properly as groupings of cooccurring syndromes has not yet been definitely demonstrated. Although the position of the Anxious syndrome among other syndromes was clear, i.e., opposite to externalizing behaviors, this was much less the case with the Withdrawn/Depressed scale. The Withdrawn/Depressed scale constructed in this study had only moderate loadings on the Internalizing gouping, but also relatively high cross-correlations with the Externalizing grouping. This issue is further illustrated in the American study reported by Achenbach (1992), where the Withdrawn scale had equal loadings both on the Internalizing and Externalizing grouping. As Campbell (1990) remarks, it may be that internalizing problems are more global in early childhood and become more specific and focused with development. This suggestion was also supported by some of our longitudinal data. Internalizing problem scores appeared to be relatively unstable over the two-year period, while the Withdrawn/Depressed scale predicted both internalizing behaviors and behaviors not subsumed under the

internalizing or externalizing grouping. Conversely, while the Oppositional syndrome emerged as a good predictor of later externalizing problems, it also had a consistent and relatively strong contribution in the prediction of later internalizing behaviors.

The internal consistency and test-retest stability coefficients were quite acceptable for most of the CBCL/2-3 scales. Internal consistency was low for Somatic Problems, which may be expected because this syndrome was composed of very few items. The test-retest reliability of the Withdrawn/Depressed scale was moderate. Interparent agreement for the scale scores was generally moderate, but quite comparable with the level of agreement found for the CBCL/2-3 by Achenbach (1992), as well as for other cross-informant comparisons (Achenbach, McConaughy, & Howell, 1987). The interparent agreement for scales reflecting behavior in the internalizing spectrum was lower than for externalizing scales, which compared to Achenbach's (1992) findings for the 3-year age-group.

The factor structure of the CBCL/2-3 resulting from this study was fairly invariant across different Dutch samples. Correlations of scale scores using the Dutch and American syndromes showed high concordance at the level of scores across cultures for all scales, except Somatic Problems. Both the high congruency of the factor structure across different samples, and the similarity of most scales across cultures are supportive of the validity of the differentiation of young children's problem behaviors into syndromes. These syndromes reflect oppositional behavior and aggression constituting an externalizing grouping, and anxiety and withdrawal, constituting an internalizing grouping. Sleep problems represent a separate syndrome at this age.

Unexpectedly, we also found a distinct Overactive syndrome. The validity of this syndrome was supported by the results of the follow-up study, in which the Overactive score had a large independent contribution to the prediction of Attention Problems scores as rated by both parents and teachers. Until now, a separate syndrome including items indicative of hyperactivity and attention problems had only been found in one other study using rating scales. Behar and Stringfield (1974) obtained a distinct Hyperactive-Distractable factor using a teacher rated questionnaire. However, only one-sixth of their sample was 3 years old. The present study thus confirmed the possibility of a further differentiation of toddler's externalizing problem behaviors using parent ratings.

The convergence between American and Dutch syndromes emerged despite differences in factor analytic approach, as might be expected with large numbers of variables (Snook & Gorsuch, 1989). The use offactor analysis instead of principal components analysis made it possible to employ techniques that provided better structure (PROMAX) and more reliable estimates of factor loadings, that were of value in comparisons across samples (LISREL). This approach yielded highly discriminating scales, which also resulted in a lower correlation (.40) between the Internalizing and the Externalizing scale than was obtained for similar scales of the American version of the CBCL/2-3 (.75), and for the American and Dutch versions of the CBCL/4-18. However, in using a different factor-analytic technique as well as differently composed samples than those used by Achenbach (1992), the cross-cultural comparability of results was lost. In other words,

we are not able to decide to what extent the somewhat different factor structure of the CBCL/2-3 obtained in this study compared to the one found for American samples was the result of the use of different factor-analytic methods and different samples, or was a reflection of true cross-cultural differences in young proechoolers' problem behavior. An exact replication of Achenbach's (1992) analyses, however, may very well be performed using the data obtained in this study in addition to the analyses reported in this thesis.

The construct validity of the CBCL/2-3 was confirmed by its significant associations with the Richman Behavior Checklist using both continuous and categorical scores. We found a mean correlation between the BCL and the total problems score of .65, and a concordance between categorical scores of 87%. Further, most correlations between the MCDI General Development and CBCL/2-3 scores were low (with a mean of -.06), and nonsignificant. The evidence of construct validity of the CBCL/2-3 is limited by the absence of other validated measures of behavior problems for young preschoolers that use methods other than parental ratings.

The criterion-related validity of the CBCL/2-3 was supported by the significant differences between demographically matched referred and nonreferred children's scale scores, which were highly comparable to the American findings reported by Achenbach (1992). Relatively small effects of referral status were found for the Anxious and Somatic Problems scale, however. The total correct classification rate of 73.3% using the total problem score was somewhat lower than the 80.1% found for CBCL total problem scores for a sample of Dutch children aged 4-16 (Verhulst, Akkerhuis, & Althaus, 1985).

Findings on the discriminative power of the CBCL/2-3 have implications for its use as a tool to identify "deviant" children. Although the rate of correct classification of referred versus nonreferred children was reasonably adequate, it seems not warranted to consider the total problem score of the CBCL/2-3 as a valid discriminator between "disturbed" and "nondisturbed" children. Thirty-seven percent of the referred children had no deviant total problem score, while obviously the parents and caretakers of these children were sufficiently concerned about their behavior to seek professional help. Furthermore, of the children having one or more deviant syndrome scores (i.e., above the 95th percentile) only 74.4% would also be classified as deviant using only the total problem score. However, as noted before, "disturbed" and "nondisturbed" should be regarded as categories of dubious validity when applied to 2-3-year-olds, because at this age behavior may fluctuate extensively from day to day, because difficult behavior and intense mood states are so common at this age, and because psychopathological disorders in the sense of distinct diagnostic entities have hardly been identified for this age group.

Occasionally, an individual problem may indicate the presence of severe psychopathology, e.g. self-mutilation in a young child. Most isolated problem behaviors as such are not indicative of psychopathology. However, high scores on those individual items that discriminated best between referred and nonreferred children may be a sign of serious problems. The following emerged as the best discriminating items in both the Dutch and American analyses: *20. Disobedient*, *25. Doesn't get along with other kids*,

29. Easily frustrated, *47. Nervous*, *58. Punishment doesn't change his/her behavior*, *82. Sudden changes in mood or feelings*, *85. Temper tantrums or hot temper*, *88. Uncooperative*, and *96. Wants a lot of attention*. However, given the generally low reliability of single items, high scores on separate items should not be regarded as anything more than possible signs of psychopathology.

9.2 Problem behavior of preschoolers in the community

The prevalence rates of parent-reported problems and the scale scores obtained on children in the community sample constitute a data base that complements the population based information on parent-reported problem behavior in Dutch children aged 4-16, that was presented by Verhulst, Akkerhuis, and Althaus (1985). These data may be used as a yardstick against which CBCL/2-3 scores of other individuals and groups of children ages 2-3 can be compared. Although children from urban areas were somewhat underrepresented in our sample it may be considered fairly representative of the general population, given the response rate of more than 90%.

The prevalence rate of individual problem items ranged from 1% to nearly 80%. Sixteen items were reported for fifty percent or more of the children in this sample, and may be regarded as typical for the age group studied. Most of these items reflect oppositional/defiant behaviors and limited emotional and behavioral control.

We found a precentage of 12.6% of children having one or more deviant syndrome scores at this age. This figure is highly comparable with prevalence figures reported in the United States, Canada, and Great Brittain. However, the meaning of prevalence figures as these remains unclear. Given the absence of an independent standard for deviancy at this age (and other ages as well), the problem of obtaining prevalence estimates reflecting true morbidity cannot be solved, even when a two-stage approach (Dohrenwend & Dohrenwend, 1982) to case-finding would be applied.

With the exception of Anxious/Depressed and Sleep Problems, mean scores of Dutch and American children on the American (1992) scales were the same. This shows that when the American scales are used, for instance to enhance the cross-cultural comparability of studies, comparison of the obtained data to Dutch or American normative bases would yield very similar results.

In this community sample, parental education, socioeconomic position of the family, health of the child, parental mental health, and maternal parenting attitude were the strongest correlates of young preschoolers' problem behavior as assessed by the CBCL/2-3. Factors were selectively related to syndrome scores. Internalizing problems were specifically related to low parental education, ill health of the child, and poor mental health in both parents. Externalizing behaviors were related to low socioeconomic status of the family, ill health of the child, mother being exhausted, irritated, and physical punishing the child, and the presence of stressful life-events. Sleep problems were specifically related to ill health of the child, number of days in hospital, poor maternal health, and marital conflicts. Somatic problems held a specific relation to maternal

physical and mental health.

We may conclude that the findings on the relation of important environmental and child factors to CBCL/2-3-scores in this study were in congruence with findings from other epidemiological studies using other measures of problem behavior with similar age groups. It is remarkable that all relations between sociodemographic and family factors, found in the London study by Richman, Stevenson, and Graham (1982), were replicated in the present community sample.

9.3 Problem behavior of referred preschoolers

To our knowledge, the data on the sample of 458 clinically referred preschoolers reported in this thesis constitute a unique overview of referral characteristics and parent reported problems at the young age of 2-3 years. As always, the representativeness of the referred sample, compared to all children in the community who have significant problems, or compared to all children who are referred to mental helath services at this age is unknown. We have provided data from a broad range of institutions in the Netherlands that provide services to young preschoolers. Known problems with research on clinical samples such as high comorbidity in the children and referral bias may have existed to their full extent in this sample.

Mean CBCL/2-3 scale scores were in this referred sample on the average almost two times as high as in the community sample. Almost 70% of the children had one or more deviant syndrome scores. Unlike in the community sample, deviant scale scores concurred with deviant total problem scores in at least 80% of all cases. This may be regarded as an indication of high comorbidity in this clinical sample. The large difference between boys and girls on proportion of deviant Aggressive scores found in the community sample was not present in this sample. Apparently, once children at this age are referred for professional help levels and types of problem behavior may be expected to be not dramatically different for boys and girls.

In this sample, several problem items were scored as "very true" of "often true" for 40% or more of the children. These were: *5. Can't concentrate*; *6. Can't sit still*; *8. Can't wait*; *15. Defiant*; *16. Demands must be met*; *20. Disobedient*; *44. Angry moods*; *58. Punishment doesn't change behavior*; *59. Quickly shifts activity*; *66. Screams*; *81. Stubborn*; *85. Temper tantrums*; *96. Wants attention*. Each of these items describes behavior indicative of oppositional and overactive behavior. This is confirmative of the wide-spread belief that disturbing and annoying child behavior may be a prime reason for parents to seek help for their children. It was also striking that, compared to older age groups, symptoms of depression were reported for very few children, even in the clinical sample. Again, this may be an indication that internalizing psychopathology is still fairly underdeveloped at this age. That is, although withdrawn behavior may be fairly conspicuous at this age, depression is not.

The majority of the children in the clinical sample presented with conduct problems

(56%), developmental problems (35%), and/or management problems (27%). Emotional, sleep, contact, and family problems were each presented by 12-14% of the children. Temper tantrums, eating problems, and toileting problems were each presented by less than 10% of the children. Analyses of differences between services showed that the proportion of referral complaints regarding developmental problems was higher in more specialized services, whereas the proportion of management problems was higher in less specialized services. This may indicate that at this age developmental problems constitute a more serious problem than management problems, which may be more amenable to change.

CBCL/2-3 scales were also differentially and in predictable directions related to referral complaints and ICD-10 diagnoses, based on file information. This is an important finding because it relates parent reports on problem behavior to information on problems obtained from independent sources. The associations between complaints and diagnoses and syndrome scores found in this study therefore further corroborate the contruct validity of the CBCL/2-3 syndromes.

Some might have expected (or hoped) to find clear differences between parent rated problem scores for children with Pervasive Developmental Disorder (PDD; one of the few relatively well-described child psychiatric disorders at this age) and children with other diagnoses. Although differences were found indicating that children with PDD had lower Oppositional, Aggressive, and Sleep Problems scores, but higher Withdrawn/Depressed scores than children receiving other diagnoses, this pattern of differences was not unique, since it was also found for the mentally retarded/low intelligence group. It is clear that the CBCL/2-3 should not be regarded as a screening instrument for autism. More specialized instruments exist to that end.

Comparison of sex, age, and SES effects in the community and clinical sample shows consistencies and inconsistencies. Consistent sex effects all regarded aggressive behavior: in both samples boys showed more destructive behavior, more fighting, hitting and attacking, and less sensitivity to punishment than girls. Age effects that were consistent across samples mainly regarded items reflecting withdrawn, depressed, and anxious behavior. SES differences that reappeared in both samples were more miscellaneous, and mainly reflected dependency, demandingness, and boisterous behavior.

Two cautionary remarks regarding the sex, age, and SES differences found in these studies are appropriate here. First, all sex, age, and SES differences in problem scores on both the individual item and scale score level were relatively small. Second, whether these differences reflect true differences in frequency and intensity of problem behavior in children cannot be readily ascertained, because the scores may be influenced by the child's behavior as well as by the parents' perception and readiness to report.

Comparison of the findings on correlates of problem behavior in the clinical and community sample shows that indices of educational/occupational level of the parents, indices of child health or hospitalization, parental mental health, caregiving, and single parenting were related to the same problem scales in both samples. Consistent correlates

of the broad-band Internalizing scale were parental education, child health and parental mental health. Both Externalizing and total problem scores showed consistent relations to parental education, child health and caregiving history.

Replication of findings on correlates of parent reported problems in the community and clinical sample is a strong indication for the validity of these relations, because information on correlates was obtained in different ways in both samples.

Using the same CBCL/2-3 syndrome scales in a sample of 1306 monozygotic and dizyotic twin pairs, Van Den Oord (1993) found that at average 65 percent of the variance of the scales (excluding Somatic Problems) could be attributed to genetic influence and 33% percent to environmental influences. The environmental influences may be of a shared (12%) or nonshared (21%) origin. Children living in the same family may share the same or similar environmental experiences (shared environment), or be influenced by environmental influences that are specific for that individual (nonshared environment). From that study, however, it could not be concluded which specific environmental factors might be at work. The variables that were consistent correlates of problem behavior in both the community and clinical sample may provide some insight as to the kind of factors responsible for the two types of environmental influence. Birth order, pre-, peri-, and postnatal circumstances, health of the child currently and in the past, maternal parenting attitudes and behaviors specific for one child, and stressful life-events directly affecting one child may account for proportions of *nonshared* environmental variance. Correlates subsumed under the family composition and social position heading, parental health, parental psychopathology, parental generalized child-rearing attitudes and behaviors, and stressful life-events affecting all children in the family may account for proportions of *shared* environmental variance. As might be expected from the analyses reported by Van Den Oord (1993), proportions of variance in the present study's Time 2 scale scores accounted for by environmental variables over and above early ratings of child behavior were small indeed.

9.4 Limitations of the prevalence studies

One limitation of the studies discussed above is that the influence of several potentially important correlates that probably are active during the first and second year of life was not assessed. For example, several studies found that difficult temperament traits during infancy was predictive of problem behavior at preschool age (Bates, Maslin, & Frankl, 1985; Lee & Bates, 1985), most notably conflicted mother-child encounters, aggression, and noncompliance. Also, the quality of the attachment relationship between the infant and the primary caregiver seems to have some impact on the occurrence of problem behaviors in preschool and school (Lewis et al., 1984; Sroufe, 1983), especially when combined with enduring suboptimal parenting skills (Erickson, Sroufe, & Egeland, 1985). As noted earlier, the study's main aims regarded assessing the validity of the CBCL/2-3. One of the means to show this was to relate scale scores to variables which could be

regarded reasonable candidates to be correlated factors of problem behavior at this age. Thus, we had no plan whatsoever to include all candidate *predictors* of problem behavior in the study. To study the etiology of problem behaviors in young preschoolers, we need longitudinal birth cohort studies using large samples, and including sensitive measures of social, child, and family factors. In these studies, the CBCL/2-3 may play a significant role as a sensitive and differentiated outcome measure.

A more important limitation of both cross-sectional studies is the fact that direct information on problem behavior was obtained from only one informant, mostly the mother. These studies can therefore be regarded studies of maternal reports of problem behavior and their correlates. Using only one informant we are never sure that existing problems in a child are veridically reported. For example, a mother who reports anxiety in her child may do this because the child displays more signs of anxiety, or because she is more sensitive to these clues.

Obtaining information from more than one informant is important because different informants tend to differ in ratings of behavioral and emotional problems (Achenbach, McConaughy, & Howell, 1987), although each informant may provide reliable data. This may have several reasons.

First, an informant may simply not be well informed about the child's behavior, due to frequent absence or because he/she is not interested in the child.

Second, a child may show different behaviors in the presence of different informants. This may be so because the situation in which the target child and the informant meet (e.g., during routine versus play activities) may elicit certain behaviors in the child, or because one informant may elicit other behavior in the child than another. For example, caretakers who hold the belief that they can control the behavior of the child are effective in eliciting less difficult child behavior from generally "difficult" children than those who hold the belief that this behavior is beyond their control (e.g., see Bugenthal and Shennum, 1984).

Third, different informants may be biased seriously in their reports. For example, in situations where, in the view of the informant, the level of reported problems may seem influential regarding decisions affecting the child or the family or both, informants may exaggerate or deny problems deliberately. *In*deliberate distortion of reports may be due to bias in interpreting the child's behavior, current mood, inappropriate standards for behavior, etc. Thus, informants may paint a different picture of a given child despite the presence of the same behavioral information.

A further important limitation of this study may be that information on correlates of problem behavior in the community sample was derived from the same informant as the information on problem behavior. Family and child factors that appeared to be related to problem behavior in the child may be influenced in unknown ways by parental characteristics also influencing problem ratings. However, in the clinical sample the information on correlates was obtained by a truly independent method. Therefore, we may have

confidence in findings on relations between factors and reported problem behavior that were replicated in both samples.

Ideally, we should have applied multiple methods of data collection to assess each variable of interest including life-data, observational methods, reports from both parents, and standardized testing. However, as shown in Chapter 2, standardized methods to obtain relevant and reliable data on children in the age range studied are hardly available. Some methods to assess correlates could and should have been more rigourous, however, and could, for example, have included medical reports on child health or standardized scales for measuring parental psychopathology.

9.5 Continuity and change

Parent ratings of CBCL/2-3 scales significantly predicted parent ratings of CBCL/4-18 scales at follow-up two years later. All Time 1 scales except Sleep Problems and Somatic Problems accounted for medium amounts of variance in Time 2 scale scores. Cross-time correlations of Externalizing and total problem scores were large, and showed significantly larger stability than Internalizing scores. These findings were supported by categorical analyses. Variance in seven out of eight CBCL/4-18 syndromes at follow-up was accounted for by CBCL/2-3 syndromes over and above the contribution of family characteristics and other child characteristics measured at initial and follow-up assessment.

Stabilities found for parent ratings, however, indicated general rather than scale specific stability, which is in contrast to reports of findings on older children. Most notably, the Oppositional and Withdrawn/Depressed scale functioned as very broad predictors. Strong oppositional behaviors at age 2-3 tended to be predictive of both internalizing and externalizing problems two years later. Withdrawn/depressed behaviors at this young age tended to predict internalizing problems as well as problems covered by the syndromes that are not strong representatives of either of the broad-band groupings. One highly specific predictor was the Overactive scale showing strong relations with the Attention Problems syndrome. Although the CBCL/2-3 Aggressive and Anxious scales showed predictive relations in expectable directions, they were in no case the strongest direct predictor of any CBCL/4-18 syndrome.

While the stability of the parent ratings of problem behavior from age 2-3 to age 4-5 was considerable, the predictive value of these parent ratings regarding teacher reports of problem behavior were quite low using either quantitative of qualitative information. The strongest correlates of teacher ratings of children's problem behavior were teacher ratings of school competence as measured by task-related behavior, social behavior, affect, and self-help skills. Apparently, a large part of the variance in preschool teachers' ratings of pupils' behavioral/emotional problems is accounted for by the perceived level of child competence regarding school tasks.

We may conclude that although several narrow-band syndromes can reliably and validly be distinguished in 2-3-year-old children's problem behavior as reported by pa-

rents, these syndromes are of limited value in predicting specific narrow-band syndromes two years later. However, the narrow-band dimensions distinguished at early preschool age clearly are predictive of specific broad-band dimensions at Kindergarten age. Predictive validity of the CBCL/2-3 defined in this way was high compared to other findings from longitudinal studies on both preschoolers and older children.

What is the meaning of low specific stability of syndromes from the prespective of construct validation? One may speculate that several syndromes identified at age 2-3 may be expressions of ongoing conflicts between parents and young preschoolers (e.g., the Oppositional syndrome) or of early signs of disturbance (e.g., Withdrawn/Depressed) which, if not resolved of alleviated, tend to develop into a broad range of behavioral/emotional problems. In the case of oppositional behavior this development might be influenced by parental reactions to the behaviors, and end up in three different directions. For example, the frequently reported "clash of wills" between young preschoolers and their parents, of which behaviors subsumed in the Oppositional syndrome may be an expression, might be ended in a way that is satisfying for both parties. This, in due time, would result in lower overall child problem scores. Other parent-child pairs might keep their conflicts which might result in high levels of parent-reported *externalizing* problems. Conversely, parents who tend to consistently solve the conflicts with too little respect for the child's autonomy, e.g. by frequent threats and punishment, might eventually produce anxious and withdrawn behaviors in their child. At the group level, these three developmental pathways may be reflected in moderate correlations of early oppositional behavior scores with later signs of both externalizing and internalizing problems. In contrast, some aggressive and overactive behaviors at young preschool age may be sufficiently specific as to persist if present at a high initial level. To further study continuities and changes of oppositional and withdrawn / depressed behaviors, young preschoolers might be selected on basis of their CBCL/2-3 scores, and be intensively and repeatedly followed during their development using diverse methods and informants.

An interesting, but incidental finding was that some CBCL/4-18 and TRF scale scores were higher in the follow-up of the community sample than those reported for a comparable of 4-5-year olds assessed eight years earlier by Verhulst et al. (1985). Comparisons showed that in our sample of 4-5-year-olds assessed in 1991 the CBCL Withdrawn, Somatic Complaints, Thought Problems, and Internalizing scores, and the TRF Somatic Complaints, Anxious/Depressed, and Internalizing scores were higher than in the Verhulst et al. sample. These findings may be indicative of a small trend of increasing internalizing problems in the general population of young preschoolers, independent of increases in the overall level of reported problems.

9.6 Is the CBCL/2-3 a useful instrument?

Use of the CBCL/2-3 in the diagnostic process

Given its demonstrated reliability and validity, the CBCL/2-3 may provide a good alternative to the only other well-validated parent-rating scale for 3-year-olds, the BSQ (Richman, Stevenson, & Graham, 1982). Especially when a detailed view of the young preschooler's behavior is desired, the large item pool and the differentiated structure of the CBCL/2-3 may be welcomed. Achenbach in his *Manual for the Child Behavior Checklist/2-3 and 1992 Profile* provides a large amount of suggestions for the practical application of the CBCL/2-3.

When judging whether a checklist is a useful tool to assess behavioral and emotional problems of referred children (and youth) one should regard several characteristics of the instrument apart from its psychometric properties. Relevant features are (1) the range of child characteristcs that are covered by the instrument; (2) the position the instrument can take in the process of diagnosis and classification; and (3) the value that may be attached to the information obtained by the instrument when used with individual children.

To achieve a valid assessment of the type and severity of the child's problems we need information on different domains of functioning including personality and psychopathology, developmental disorders, cognitive development, physical condition, and psychosocial conditions. The information obtained with the CBCL/2-3 pertains to the first domain. Furthermore, the CBCL/2-3 is designed to assess relatively common behavioral and emotional problems in young preschoolers. We may say that the CBCL/2-3 can be used as a broad-band instrument that gives a valid first impression of the type and severity of psychopathology in the 2-3-year-old child. Beside the CBCL/2-3 other, more specific instruments may be used to cover more specific psychopathology (e.g., autism), and and to cover other domains of functioning.

Further, we should indicate which part of the diagnostic and treatment process is covered by the instrument. At least five phases of the diagnostic and treatment process may be discerned in which information on the child is processed. These are: (1) referral; (2) diagnostic investigation; (3) integration of information to achieve diagnosis and choice of treatment; (4) treatment; and (5) evaluation of treatment results. In each of these phases the diagnostician or therapist needs to gather, integrate and communicate information about the child and its environment. The CBCL/2-3 may prove to be useful in all stages in which a standardized method to obtain information on the child's behavior from parents or caretakers is desirable. This may be especially the case at stages 1 and 5 indicated above. The CBCL/2-3 covers several aspects of deviancy that can be compared to sex-appropriate norms: frequency, intensity, and constellation of behavioral/emotional problems. Other parameters of deviancy such as chronicity of the problems and the (social) context in which they occur are not covered by the instrument, but should be explored through direct contact with diverse informants.

As indicated earlier, informants who see the child in different contexts and who have

different relationships with him/her often differ greatly in their descriptions of the presence and seriousness of problem behavior and the circumstances in which these occur. Discrepancy between information obtained from different informants should not be equalled to unreliability, since each informant can give reliable though different information. However, although good psychometric properties of an instrument are a prerequisite to its proper use, they are no cure against apparent unreliability of an individual informant. When interpreting results obtained with the CBCL/2-3 an assessment of the reliability of the individual informant should always be included by means of interview and observation.

Use of the CBCL/2-3 as a screening instrument

We might imagine that, if a child could be identified as possibly disturbed at two or three years we might prevent problems arising at school by taking appropriate measures after further inquiry of the extent and nature of the problems. In our opinion, routine screening by any service, such as family doctors, consultation agencies, or daycare facilities may only be justifiable if a disease or disorder is very well described, when non-detection has serious consequences, when the screening efficiency of the instrument to be used is very high, and when effective help can be offered. From this perspective screening of developmental or behavioral/emotional problems remains a risky enterprise.

At the age of 2-3 years perceived problem behaviors may reflect age-appropriate behaviors or difficult transitions from one stage of development to the next. Behavioral difficulties reported to family doctors or outpatient mental health professionals may often reflect misunderstandings on the part of parents about the normal course of development and the particular tasks of a developmental stage. Further, common developmental problems may often be exacerbated by parental mismanagement, family stress, or combinations thereof. The behavioral difficulties may thus prove to be "transient" after parents have received information on the relative normality of the behavior, when stress has dissipated, or when parents have received information on how to deal with certain difficult behaviors.

In addition to these considerations, we should regard the screening efficiency of the CBCL/2-3. By far not all children who need help according to their parents and others or who will have adjustment problems at Kindergarten age could be detected or predicted from knowledge of their behavior at age 2-3 as assessed by the CBCL/2-3. The CBCL/2-3 total problem score showed a sensitivity of 80% to discriminate between referred and nonreferred 2-3-year-olds, and a predictive sensitivity of 63% regarding parent-reported problems at age 4-5, while the concurrent and predictive specificity were 69% and 84%. This resulted in an overall efficiency of 73%, both concurrently and predictively. Although these figures may seem to reflect adequate screening power, the numbers of false positives and false negatives produced by the instrument would be so high as to prohibit its use as the sole basis of decisions regarding referral or intervention.

Richman et al. (1982) were confronted with a similar situation exploring the possibilities for screening by means of the BSQ and BCL. These authors advised that

instead of screening, continual monitoring of the children by parents and teachers is necessary if disturbed children in need of help are to be identified and appropriate measures taken. In our view, this is a very reasonable position. If parents or others signal behavioral/emotional problems to health workers, family doctors or other professionals, the CBCL/2-3 could very well be used as an initial description and normative comparison of the problems. The CBCL/2-3 could then be used as a starting point for further exploration of specific problems before a decision is made whether the child or his/her parents need further help, or merely careful monitoring is sufficient.

Use of the CBCL/2-3 as a research tool

Research on the epidemiology, diagnosis, etiology, and treatment of behavioral/emotional problems in young preschoolers is still in its infancy. The studies reported in this thesis represent one of the efforts in this area. Achenbach (1992) gives an extremely valuable overview of the possibilities of the CBCL/2-3 for research use, which need no repetition. Further, the yearly updated *Bibliography of published studies using the Child Behavior Checklist and related materials* (Brown & Achenbach, 1993), provides an overview of the more than 1000 published studies in which the CBCL and its younger brothers and sisters were used. In the Netherlands, the Dutch version of the CBCL/2-3 was already used in a genetic study of problem behavior of 3-year-old twins (Van Den Oord, 1993), and in a study of behavior problems in very low-birthweight children at age 3 (Weisglas et al., 1993).

We would like to suggest three areas of future research that would be of interest to advance the standardized assessment of behavioral/emotional problems of young preschoolers.

A first need regards the development and standardization of measures equivalent to the CBCL/2-3 to obtain information from different informants, such as workers in daycare facilities. To maintain high comparability with parent ratings, the CBCL/2-3 items could serve as a basis for a rating form to be used in daycare, and subsequently normative data could be obtained for children visiting non-therapeutic daycare facilities. The development of a direct observation form including comparable items would be highly useful to perform further validation studies.

Second, a study of systematic influences on informants' ratings of problem behavior in children would be of considerable value in understanding apparent differences between mothers', fathers', and preschool teachers' reports. Using maternal, teacher, and child ratings of conduct disorder, Fergusson and Horwood (1987) estimated that only between 28% and 40% of the variance in maternal and teacher ratings was ascribable to variations in the child's generalized behavioral tendencies, and the remaining variance either to method-specific factors or random errors of measurement. Although examples of elegant studies on this topic exist (e.g., Brody & Forehand, 1986; Mills & Rubin, 1990), this area of research needs much more attention.

Third, although the results from our study on the clinical sample provide some evidence of meaningful relations between diagnostic information and parent ratings of

problem behavior on CBCL/2-3 scales, the relation between these scales and results of more direct clinical and observational assessment of referred children needs to be investigated. Outpatient units of university child psychiatry departments would seem excellent sites to accomplished this research need.

In this chapter, we discussed findings from studies on the value of the CBCL/2-3 as a parent rating scale of problem behavior in children ages 2-3 years. In our appraisal, we have provided substantial evidence that the CBCL/2-3 is a psychometrically sound instrument yielding a differentiated picture of toddlers' behavioral and emotional problems. We have constructed scales for the scoring of young children's behavior; provided psychometric and normative data for the instrument; provided evidence of theoretically plausible associations of the instrument's scale scores with environmental, family, and child factors and diagnostic data, and of their association with ratings of behavioral/emotional problems of the child at school entry. However, whether the CBCL/2-3 is a useful instrument will ultimately be answered by using it in clinical practice and in research. We hope that many will do so.

References

Abidin, R.R. (1983). *Parenting stress index: Manual.* Charlottesville: Pediatric Psychology Press.

Achenbach, T.M. (1966). The classification of children's psychiatric symptoms: A factor-analytic study. *Psychological Monographs, 80* (No. 615).

Achenbach, T.M. (1978). The Child Behavior Profile: I. Boys aged 6-11. *Journal of Consulting and Clinical Psychology, 46,* 478-488.

Achenbach, T.M. (1985). *Assessment and taxonomy of child and adolescent psychopathology.* Beverly Hills, CA: Sage.

Achenbach, T.M. (1990). What is "developmental" about developmental psychopathology? In J. Rolf, A.S. Masten, K.H. Nuechterlein, & S. Weintraub (Eds.), *Risk and protective factors in the development of psychopathology.* Cambridge: Cambridge University Press.

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

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

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

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

Achenbach, T.M. (1992). *Manual for the Child Behavior Checklist/2-3 and 1992 Profile.* Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T.M. (1993). *Empirically based taxonomy: How to use syndromes and profile types derived from the CBCL, TRF, and YSR.* Burlington, VT: University of Vermont, Department of Psychiatry.

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

Achenbach, T.M., & Edelbrock, C.S. (1978). The classification of child psychopathology: A review and analysis of empirical efforts. *Psychological Bulletin, 85,* 1275-1301.

Achenbach, T.M., & Edelbrock, C.S. (1981). Behavioral problems and competencies reported by parents of normal and disturbed children aged four through sixteen. *Monographs of the Society for Research in Child Development, 46* (Serial No. 188).

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

Achenbach, T.M., & Edelbrock, C. (1986). *Manual for the Teacher's Report Form and Teacher Version of the Child Behavior Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

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

Achenbach, T.M., Edelbrock, C., & Howell, C.T. (1987). Empirically based assessment of the behavioral/emotional problems of 2- and 3-year old children. *Journal of Abnormal Child Psychology, 15*, 629-650.

Achenbach, T.M., McConaughy, S.H., & Howell, C.T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*, 213-232.

Achenbach, T.M., Verhulst, F.C., Baron, G.D., & Althaus, M. (1987). A comparison of syndromes derived from the Child Behavior Checklist for American and Dutch boys aged 6-11 and 12-16. *Journal of Child Psychology and Psychiatry, 28*, 437-453.

American Psychiatric Association. (1987). *Diagnostic and Statistical Manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

Anastopoulos, A.D., Guevremont, D.C., Shelton, T.L., & Dupaul, G.J. (1992). Parenting stress among families of children with Attention Deficit Hyperactivity Disorder. *Journal of Abnormal Child Psychology, 20*, 503-520.

Anderson, J.C., Williams, S., McGee, R., & Silva, P. (1989). Cognitive and social correlates of DSM-III disorders in preadolescent children. *Journal of the American Academy of Child and Adolescent Psychiatry, 28*, 842-846.

Barkley, R.A. (1988). Child behavior rating scales and checklists. In M. Rutter, A.H. Tuma, & I.S. Lann (Eds.), *Assessment and diagnosis in child psychopathology* (pp. 113-155). New York: Guilford Press.

Barron, A., & Earls, F. (1984). The relationship of societal and temperament factors in behavior problems in three-year-old children: A cross-national replication. *Journal of Child Psychology and Psychiatry, 25*, 23-33.

Bates, J.E. (1980). The concept of difficult temperament. *Merril-Palmer Quarterly, 26*, 299-319.

Bates, J.E. (1987). Concepts and measures of temperament. In G.A. Kohnstamm, J.E. Bates, & M.K. Rothbart (Eds.), *Temperament in childhood* (pp. 3-26). New York: Wiley.

Bates, J.E., & Bayles, K. (1988). The role of attachment in the development of behavior problems. In J. Belsky & T. Nezworski (Eds.), *Clinical implications of attachment* (pp. 253-299). New York: Erlbaum.

Bates, J.E., Maslin, C.A., & Frankel, K.A. (1985). Attachment security, mother-child interaction and temperament as predictors of behavior problem ratings at age three years. In I. Bretherton & E. Waters (Eds.), Growing points of attachment theory and research. *Monographs of the Society for Research in Child Development, 50* (Serial No. 209, pp. 167-193).

Bax, M., & Hart, H. (1976). Health needs of preschool children. *Archives of Diseases in Childhood, 51*, 848-852.

Bayley, N. (1969). *Bayley Scales of Infant Development*. New York: Psychological Corporation.

Becker, W.C. (1960). The relationship of factors in parental ratings of self and each other to the behavior of Kindergarten children as rated by mothers, fathers, and teachers. *Journal of Consulting Psychology, 24*, 507-527.

Behar, L.B. (1977). The Preschool Behavior Questionnaire. *Journal of Abnormal Child Psychology, 5*, 265-276.

Behar, L.B., & Stringfield, S. (1974). A behavior rating scale for the preschool child. *Developmental Psychology, 10*, 601-610.

Beitchman, J.H., Wekerle, C., & Hood, J. (1987). Diagnostic continuity from preschool to middle childhood. *Journal of the American Academy of Child and Adolescent Psychiatry, 26*, 694-699.

Bell, R.Q., Waldrop, M.F., & Weller, G.M. (1972). A rating system for the assessment of hyperactive and withdrawn children in preschool samples. *American Journal of Orthopsychiatry, 42*, 23-34.

Berden, G.F.M.G. (1992). Ontwikkeling van een levensgebeurtenissenvragenlijst en een levensgebeurtenisseninterview. Unpublished doctoral dissertation. Erasmus University, Rotterdam.

Berden, G.F.M.G., Althaus, M., & Verhulst, F.C. (1990). The relation between major life events and changes in the emotional/behavioural functioning of children. *Journal of Child Psychology and Psychiatry, 31*, 949-959.

Berkson, J. (1946). Limitations of the application of fourfold table analysis to hospital data. *Biometrics Bulletin, 2*, 47-53.

Bird, H.R., Gould, M.S., Yager, T., Staghezza, B., & Canino, G. (1989). Risk factors for maladjustment in Puerto Rican children. *Journal of the American Academy of Child and Adolescent Psychiatry, 28*, 847-850.

Brody, G.H., & Forehand, R. (1986). Maternal perceptions of child maladjustment as a function of the combined influence of child behavior and maternal depression. *Journal of Consulting and Clinical Psychology, 54*, 237-240.

Brown, J.L. (1960). Prognosis from presenting symptoms of preschool children with atypical development. *American Journal of Orthopsychiatry, 30*, 383-390.

Brown, J.S., & Achenbach, T.M. (1993). *Bibliography of published studies using the Child Behavior Checklist and related materials: 1993 edition*. Burlington, VT: University of Vermont, Department of Psychiatry.

Bugenthal, D.B., & Shennum, W.A. (1984). "Difficult" children as elicitors and targets of adult communication patterns: An attributional-behavioral transactional analysis. *Monographs of the Society for Research in Child Development, 49* (Serial No. 205).

Byrne, J.M., Backman, J.E., & Smith, I.M. (1986). Developmental assessment: The clinical use and validity of parental report. *Journal of Pediatric Psychology, 11*, 549-559.

Byrne, J.M., Smith, D.J., & Backman, J.E. (1987). Cognitive impairment in preschoolers: Identification using the Personality Inventory for Children-Revised. *Journal of Abnormal Child Psychology, 15*, 239-246.

Caldwell, C., & Pianta, R.C. (1990). Developing a measure of young children's problem and competence behaviors: The Preschool Behavior Rating Scale. *Journal of Psychoeducational Assessment, 9*, 32-39.

Cameron, J. (1978). Parental treatment, children's temperament and the risk of childhood problems. 2. Initial temperament, parental attitudes, and the incidence and form of behavioral problems. *American Journal of Orthopsychiatry, 48*, 140-147.

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Campbell, S.B. (1990). *Behavior problems in preschool children: Clinical and developmental issues*. New York: The Guilford Press.

Campbell, S., Breaux, A.M., Ewing, L.J., & Szumowski, E.K. (1984). A one-year follow-up of parent-referred "hyperactive" preschoolers. *Journal of the American Academy of Child Psychiatry, 23*, 243-249.

Campbell, S.B., & Ewing, L.J. (1990). Follow-up of hard-to-manage preschoolers: Adjustment at age 9 and predictors of continuing symptoms. *Journal of Child Psychology and Psychiatry, 31*, 871-889.

Campbell, S.B., Ewing, L.J., Breaux, A.M., & Szumowski, E.K. (1986). Parent-identified behavior problem toddlers: Follow-up at school entry. *Journal of Child Psychology and Psychiatry, 27*, 473-488.

Campbell, S.B., Schleifer, M., Weiss, G., & Perlman, T. (1977). A two-year follow-up of hyperactive preschoolers. *American Journal of Orthopsychiatry, 47*, 149-162.

Campbell, S.B., Szumowski, E.K., Ewing, L.J., Gluck, D.S., & Breaux, A.M. (1982). A multidimensional assessment of parent-identified behavior problem toddlers. *Journal of Abnormal Child Psychology, 10*, 569-592.

CBS (Central Bureau of Statistics). (1987). *Standard Educational Classification 1978; Edition 1987*. Voorburg: CBS.

Cicchetti, D., & Toth, S.L. (Eds.) (1991). Internalizing and externalizing expressions of dysfunction. *Rochester Symposium on Developmental Psychopathology, Vol. II*. Hillsdale, NJ: Erlbaum.

Cohen, J.A. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). New York: Academic Press.

Cohen, S., & Bromet, E. (1992). Maternal predictors of behavioral disturbance in preschool children: A research note. *Journal of Child Psychology and Psychiatry, 33*, 941-946.

Cohen, P., Brook, J.S., Cohen, J., Velez, C.N., & Garcia, M. (1990). Common and uncommon pathways to adolescent psychopathology and problem behavior. In L.L. Robins & M. Rutter (Eds.), *Straight and devious pathways from childhood to adulthood*. Cambridge: Cambridge University Press.

Coleman, J., Wolkind, S., & Ashley, L. (1977). Symptoms of behaviour disturbance and adjustment to school. *Journal of Child Psychology and Psychiatry, 18*, 201-209.

Colligan, R.C. (1976). Prediction of kindergarten reading success from preschool report of parents. *Psychology in the Schools, 13*, 304-308.

Colligan, R.C. (1977). The Minnesota Child Development Inventory as an aid in the assessment of developmental disability. *Journal of Clinical Psychology, 33*, 162-163.

Cornely, P., & Bromet, E. (1986). Prevalence of behavior problems in three-year-old children living near Three Mile Island: A comparative analysis. *Journal of Child Psychology and Psychiatry, 27*, 489-498.

Costello, E.J. (1989). Developments in child psychiatric epidemiology. *Journal of the American Academy of Child and Adolescent Psychiatry, 28*, 836-841.

Crnic, K.A., Greenberg, M.T., Ragozin, A.S., Robinson, N.M., & Basham, R.B. (1983). Effects of stress and social support on mothers and premature and full-term infants. *Child Development, 54*, 209-217.

Crockenberg, S. (1987). Predictors and correlates of anger toward and punitive control of toddlers by adolescent mothers. *Child Development, 58*, 964-975.

Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.

Dahl, E.K., Cohen, D.J., & Provence, S. (1986). Clinical and multivariate approaches to nosology of pervasive developmental disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 25*, 170-180.

De Brock, A.J.J.L., Vermulst, A.A., & Gerris, J.R.M. (1990). Nijmeegse Ouderlijke Stress Index: Inhoud en achtergrond. Katholieke Universiteit Nijmegen, Internal publication #EP15.

De Brock, A.J.L.L., Vermulst, A.A., & Leenders, F.A.M. (1990). Huwelijkskwaliteit, ouderlijk opvoedkundig handelen en (dys)functioneren van het kind. *Tijdschrift voor Orthopedagogiek, 29*, 373-392.

De Groot, A.D. (1961). *Methodologie*. Den Haag: Mouton.

Dean, R.S., & Steffen, J.E. (1984). Direct and indirect pediatric screening measures. *Journal of Pediatric Psychology, 9*, 65-75.

Dohrenwend, B.P., & Dohrenwend, B.S. (1982). Perspectives on the past and future of psychiatric epidemiology. *American Journal of Public Health, 72*, 1271-1279.

Douglas, J. & Richman, N. (1984). *My child won't sleep*. Harmondsworth: Penguin.

Earls, F. (1980). The prevalence of behavior problems in 3-year-old children: Comparison of the reports of fathers and mothers. *Journal of the American Academy of Child Psychiatry, 19*, 439-452.

Earls, F., & Richman, N. (1980). The prevalence of behavior problems in 3-yr-old children of West Indian-born parents. *Journal of Child Psychology and Psychiatry, 21*, 99-106.

Egeland, B., Kalkoske, M., Gottesman, N., & Erickson, M.F. (1990). Preschool behavior problems: Stability and factors accounting for change. *Journal of Child Psychology and Psychiatry, 31*, 891-909.

Eisenberg, L., Landowne, J., Wilmer, M., Iamber, D. (1962). The use of teacher ratings in a mental health study: A method for measuring the effectiveness of a therapeutic nursery program. *American Journal of Public Health, 52*, 18-28.

Eisert, D.C., Spector, S., Shankaran, S., Faigenbaum, D., & Szego, E. (1980). Mothers' reports of their low birth weight infants' subsequent development on the Minnesota Child development Inventory. *Journal of Pediatric Psychology, 5,* 353-364.

Eme, R.F. (1979). Sex difference in childhood psychopathology: A review. *Psychological Bulletin, 86*, 574-595.

Erickson, M.F., Sroufe, L.A., & Egeland, R. (1985). The relationship between quality of attachment and behavior problems in preschool in a high-risk sample. In I. Bretherton & E. Waters (Eds.), Growing points of attachment theory and research. *Monographs of the Society for Research in Child Development, 50* (1-2, Serial No. 209).

Fergusson, D.M., & Horwood, L.J. (1987). The trait and method components of ratings of conduct disorder: Part II. Factors related to the trait components of conduct disorders scores. *Journal of Child Psychology and Psychiatry, 28*, 261-272.

Fischer, M., Rolf, J.E., Hasazi, J.E., & Cummings, L. (1984). Follow-up of a preschool epidemiological sample: Cross-age continuities and predictions of later adjustment with internalizing and externalizing dimensions of behavior. *Child Development, 55*, 137-150.

Fleiss, J.L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: Wiley.

Fowler, P. O., & Park, R. M. (1979). Factor structure of the preschool behavior questionnaire in a normal population. *Psychological Reports, 45*, 599-606.

Freud, A. (1965). The concept of developmental lines. *Psychoanalytic Study of the Child, 18*, 245-265.

Garber, J. (1984). Classification of childhood psychopathology: A developmental perspective. *Child Development, 55*, 30-48.

Garmezy, N. (1983). Stressors of childhood. In N. Garmezy & M. Rutter (Eds.)., *Stress, coping and development in children* (pp.43-84). New York: McGraw-Hill.

Garrison, W., & Earls, F. (1985). Change and continuity in behavior problems from the pre-school period through school entry: An analysis of mother's reports. In J.E. Stevenson (Ed.), *Recent research in developmental psychopathology*. Oxford: Pergamon Press.

Garrity, L.I., & Servos, A.B. (1978). Comparison of measures of adaptive behaviors in preschool children. *Journal of Consulting and Clinical Psychology, 46*, 288-293.

Gerbing, D.W., & Anderson, J.C. (1987). Improper solutions in the analysis of covariance structures: Their interpretability and a comparison of alternate specifications. *Psychometrika, 52*, 99-111.

Gerritsen, F.M.E.G. (1988). *VTO taalscreening 3- tot 6-jarigen: De ontwikkeling van een taalscreeningsinstrumentarium voor gebruik in de Jeugdgezondheidszorg*. Amsterdam/Lisse: Swets & Zeitlinger.

Ghodsian, M., Fogelman, K., Lambert, L., & Tibbenham, A. (1980). Changes in behaviour of a national sample of children. *British Journal of Social and Clinical Psychology, 19*, 247-256.

Goldberg, S., Corter, C., Lojkasek, M., & Minde, K. (1990). Prediction of behavior problems in 4-year-olds born prematurely. *Development and Psychopathology, 2*, 5-30.

Goldstein, H. (1987). *Multilevel models in educational and social research*. New York: Oxford University Press.

Gorsuch, R.L. (1983). *Factor analysis*. Hillsdale, NJ: Erlbaum.

Gottfried, A.W., Guerin, D., Spencer, J.E., & Meyer, C. (1983). Concurrent validity of the Minnesota Child Development Inventory in a nonclinical sample. *Journal of Consulting and Clinical Psychology, 51*, 643-644.

Gottfried, A.W., Guerin, D., Spencer, J.E., & Meyer, C. (1984). Validity of the Minnesota Child Development Inventory in screening young children's developmental status. *Journal of Pediatric Psychology, 9*, 219-229.

Goyette, C.H., Conners, C.K., & Ulrich, R.F. (1978). Normative data on Revised Conners Parent and Teacher Rating Scales. *Journal of Abnormal Child Psychology, 6*, 221-236.

Graham, P., & Rutter, M. (1973). Psychiatric disorder in the young adolescent: A follow-up study. *Proceedings of the Royal Society of Medicine, 66*, 1226-1228.

Graham, P.J., Rutter, M., & George, S. (1973). Temperamental characteristics of behavior disorders in children. *American Journal of Orthopsychiatry, 43*, 328-339.

Guérin, D., & Gottfried, A. W. (1986). Infant temperament as a predictor of preschool behavior problems. Paper presented at the International Conference on Infant Studies, Los Angeles, 1986.

Guérin, D., & Gottfried, A.W. (1987). Minnesota Child Development Inventories: Predictors of intelligence, achievement, and adaptability. *Journal of Pediatric Psychology, 12*, 595-609.

Herjanic, B., & Reich, W. (1982). Development of a structured psychiatric interview for children: Agreement between child and parent on individual symptoms. *Journal of Abnormal Child Psychology, 10*, 53-75.

Hoge, R. D., Meginbir, L., Khan, Y., & Weatherall, D. (1985). A multitrait-multimethod analysis of the preschool behavior questionnaire. *Journal of Abnormal Child Psychology, 13*, 119-127.

Holahan, C.J., & Moos, R.H. (1987). Risk, resistance, and psychological distress: A longitudinal analysis with adults and children. *Journal of Abnormal Psychology, 96*, 3-13.

Hooks, M.Y., Mayes, L.C., & Volkmar, F.R. (1988). Psychiatric disorders among preschool children. *Journal of the American Academy of Child Psychiatry, 27*, 623-627.

Hoy, E., Bill, J., & Sykes, D. (1988). Very low birthweight: A long term developmental impairment. *International Journal of Behavioural Development, 11*, 37-67.

Ireton, H., & Thwing, E.J. (1974). *Minnesota Child Development Inventory*. Minneapolis: Behavior Science Systems.

Ireton, H., Thwing, E., & Curier, S.K. (1977). Minnesota Child Development Inventory: Identification of children with developmental disorders. *Journal of Pediatric Psychology, 2*, 18-22.

Jenkins, S., Bax, M., & Hart, H. (1980). Behaviour problems in pre-school children. *Journal of Child Psychology and Psychiatry, 21*, 5-17.

Jenkins, S., Owen, C., Bax, M., & Hart, H. (1984). Continuities of common behaviour problems in preschool children. *Journal of Child Psychology and Psychiatry, 25*, 75-89.

Jöreskog, K.G.(1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36*, 109-133.

Jöreskog, K.G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika, 43*, 443-477.

Jöreskog, K.G., & Sörbom, D. (1988). *PRELIS. A program for multivariate data screening and data summarisation. A preprocessor for LISREL*. Mooresville, IND: Scientific Software.

Jöreskog, K.G., & Sörbom, D. (1989). *LISREL 7: A guide to the program and applications*. Chicago: SPSS Inc.

Kastrup, M. (1977). Urban-rural differences in 6-year olds. In P.J. Graham (Ed.), *Epidemiological approaches to child psychiatry* (pp. 181-194). London: Academic Press.

Kenny, T.J., Hebel, J.R., Sexton, M.J., & Fox, N.L. (1987). Developmental screening using parent report. *Developmental and Behavioral Pediatrics, 8*, 8-11.

Kohn, M. (1977). *Social competence, symptoms and underachievement in childhood: A longitudinal perspective*. Washington, DC: Winston.

Kohn, M., & Rosman, B.L. (1972). A social competence and symptom checklist for the preschool child: Factor dimensions, their cross-instrument generality, and longitudinal persistence. *Developmental Psychology, 6*, 430-444.

Kohn, M., & Rosman, B.L. (1973). A two-factor model of emotional disturbance in the young child: Validity and screening efficiency. *Journal of Child Psychology and Psychiatry, 14*, 31-56.

Koot, H. M., & Verhulst, F. C. (1991). Prevalence of problem behavior in Dutch children aged 2-3. *Acta Psychiatrica Scandinavica, 83* (Suppl. 367, pp. 1-37).

Koot, H.M., & Verhulst, F.C. (1992). Prediction of children's referral to mental health and special education services from earlier adjustment. *Journal of Child Psychology and Psychiatry, 33*, 717-729.

LaClave, L.J., & Campbell, J.L. (1986). Psychiatric intervention in children: Sex differences in referral rates. *Journal of the American Academy of Child Psychiatry, 24*, 430-432.

Landelijke Commissie Vroegtijdige Onderkenning Ontwikkelingsstoornissen (1988). *Een beweging in ontwikkeling. Slotadvies.* Rapport. Rijswijk.

Landis, J.R., & Koch, GG. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Larson, C.P., Pless, I.B., & Miettinen, O. (1988). Preschool behavior disorders: Their prevalence in relation to determinants. *Journal of Pediatrics, 113*, 278-285.

Lee, B.J. (1987). Multidisciplinary evaluation of preschool children and its demography in a military psychiatric clinic. *Journal of the American Academy of Child Psychiatry, 26*, 313-316.

Lee, C., & Bates, J. (1985). Mother-child interaction at age two years and perceived difficult temperament. *Child Development, 56*, 1314-1326.

Lewis, M., Feiring, C., McGuffog, C., & Jaskir, J. (1984). Predicting psychopathology in six-year-olds from early social relations. *Child Development, 55*, 123-136.

Long, A. (1941). Parents' reports of undesirable behavior in children. *Child Development, 12*, 43-61.

Luk, S.L., Leung, P.W.H., Bacon-Shone, J., & Lieh-Mak, F. (1991). The structure and prevalence of behavioral problems in Hong-Kong preschool children. *Journal of Abnormal Child Psychology, 19*, 219-232.

MacFarlane, J.W., Allen, L., & Honzik, M.P (1962). *A developmental study of behavior problems of normal children between twenty-one months and fourteen years.*. Berkeley, CA: University of California Press.

Mahler, M.S., Pine, F., & Bergman, A. (1975). *The psychological birth of the human infant*. New York: Basic Books.

Matheny, A.P. (1989). Temperament and cognition: Relations between temperament and test scores. In G.A. Kohnstamm, J.E. Bates, & M.K. Rothbart (Eds.), *Temperament in childhood* (pp. 263-282). London: Wiley.

Maziade, M., Capéraà, P., Laplante, B., Boudreault, M., Thivierge, J., Côté, R., & Boutin, P. (1985). Value of difficult temperament among 7-year-olds in the general population for predicting psychiatric diagnosis at age 12. *American Journal of Psychiatry*, *142*, 943-946.

Maziade, M., Côté, R., Thivierge, J., Boutin, P., & Bernier, H. (1989). Significance of extreme temperament in infancy for clinical status in preschool years. I: Value of extreme temperament at 4-8 months for predicting diagnosis at 4.7 years. *British Journal of Psychiatry*, *154*, 535-543.

McCarthy, D. (1972). *McCarthy Scales of Children's Abilities*. New York: Psychological Corporation.

McConaughy, S.H., Stanger, C., & Achenbach, T.M. (1992). Three-year course of behavioral/emotional problems in a national sample of 4-16-year-olds: I. Agreement among informants. *Journal of the American Academy of Child and Adolescent Psychiatry*, *31*, 932-940.

McGee, R., Partridge, F., Williams, & Silva, P. (1991). A twelve-year follow-up of preschool hyperactive children. *Journal of the American Academy of Child and Adolescent Psychiatry*, *30*, 224-232.

McGee, R., Silva, P.A., & Williams, S. (1984). Perinatal, environmental and develop mental characteristics of seven-year-old children with stable behaviour problems. *Journal of Child Psychology and Psychiatry*, *25*, 573-586.

McGee, R., Williams, S., Bradshaw J., Chapel, J., Robins, A., & Silva, P. (1985). The Rutter Scale for completion by teachers: Factor structure and relationships with cognitive abilities and family adversity for a sample of New Zealand children. *Journal of Child Psychology and Psychiatry*, *26*, 727-739.

McGuire, J., & Richman, N. (1986). Screening for behaviour problems in nurseries: The reliability and validity of the Preschool Behavior Checklist. *Journal of Child Psychology and Psychiatry*, *24*, 273-280.

McLoyd, V.C. (1990). The impact of economic hardship on black families and children: Psychological distress, parenting, and socioemotional development. *Child Development*, *61*, 311-346.

Mills, R.S.L., & Rubin, K.H. (1990). Parental beliefs about problematic social behaviors in early childhood. *Child Development*, *61*, 138-151.

Minde, R., & Minde, R. (1977). Behavioural screening of pre-school children: A new approach to mental health? In P.J. Graham (Ed.), *Epidemiological approaches to child psychiatry* (pp. 139-164). London: Academic Press.

Morris, R.J., Bergan, J.R., & Fulginiti, J.V. (1991). Structural equation modeling in clinical assessment: Research with children. *Journal of Consulting and Clinical Psychology*, *59*, 371-379.

Offord, D.R., Boyle, M.H., & Racine, Y. (1989). Ontario Child Health Study: Correlates of disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *28*, 856-860.

Olweus, D. (1979). Stability of aggressive reaction patterns in males: A review. *Psychological Bulletin, 86*, 852-875.

Parke, R. D., & Slaby, R. G. (1983). The development of aggression. In E. M. Hetherington (Ed.), Socialization, personality, and social development (Vol. 4, 547-642). In P. Mussen (Series Ed.), *Handbook of Child Psychology*. New York: Wiley.

Pianta, R. (1987). Early Behavior Rating Scale. Unpublished Document. University of Virginia, Charlottesville.

Prior, M., Leonard, A., & Wood, G. (1983). A comparison study of preschool children diagnosed as hyperactive. *Journal of Pediatric Psychology, 8*, 191-207.

Quay, H.C. (1986). Classification. In H.C. Quay & J.S. Werry (Eds.), *Psychopathological disorders of childhood* (3rd ed., pp.1-42). New York: Wiley.

Quay, H.C., & Peterson, D.R. (1983). *Interim manual for the Revised Behavior Problem Checklist*. Unpublished manuscript, University of Miami.

Reinherz, H., & Gracey, C. (1983). *Manual for the Simmons Behavior Checklist*. Simmons College School of Social Work, Boston.

Rescorla, L.A. (1986). Preschool psychiatric disorders: Diagnostic classification and symptom patterns. *Journal of the American Academy of Child Psychiatry, 25*, 162-169.

Richman, N. (1977). Is a behaviour checklist for preschool children useful? In P.J. Graham (Ed.), *Epidemiological approaches to child psychiatry* (pp. 125-137). London: Academic Press.

Richman, N., & Graham, P.J. (1971). A behavioural screening questionnaire for use with three-year-old children: Preliminary findings. *Journal of Child Psychology and Psychiatry, 12*, 5-33.

Richman, N. & Lansdown, R. (Eds.). (1988). *Problems of preschool children*. Chichester: Wiley.

Richman, N., Stevenson, J.E., & Graham, P.J. (1975). Prevalence of behaviour problems in 3-year-old children: An epidemiological study in a London borough. *Journal of Child Psychology and Psychiatry, 16*, 277-287.

Richman, N., Stevenson, J., & Graham, P.J. (1982). *Pre-school to school: A behavioural study*. London: Academic Press.

Rose, S.A., Feldman, J.F., Rose, S.L., Wallace, I.F., & McCarton, C. (1992). Behavior problems at 3 and 6 years: Prevalence and continuity in full-terms and preterms. *Development and Psychopathology, 4*, 361-374.

Rose, S.L., Rose, S.A., & Feldman, J.F. (1989). Stability of behavior problems in very young children. *Development and Psychopathology, 1*, 5-19.

Rost, H. (1992). Nijmeegse Observatieschalen voor Kleuters. Unpublished Report. Katholieke Universiteit, Nijmegen.

Rothbart, M.K., & Goldsmith, H.H. (1985). Three approaches to the study of infant temperament. *Developmental Review, 5*, 237-260.

Rubin, K.H., & Clark, M.L. (1983). Preschool teacher ratings of behavioral problems: Observational, sociometric, and social-cognitive correlates. *Journal of Abnormal Child Psychology, 11*, 273-285.

Rubin, K.H., Hymel, S., Mills, R.S.L., Rose-Krasnor, L. (1991). Conceptualizing different developmental pathways to and from social isolation in childhood. In D. Cicchetti, & S. L. Toth, (Eds.), *Rochester Symposium on Developmental Psychopathology, Vol. II: Internalizing and externalizing expressions of dysfunction* (pp. 91-122). Hillsdale, NJ: Erlbaum.

Rutter, M. (1989). Isle of Wight revisited: Twenty-five years of child psychiatric epidemiology. *Journal of the American Academy of Child and Adolescent Psychiatry, 28*, 633-653.

Rutter, M., Tuma, A.H., & Lann, I.S. (Eds.) (1988). *Assessment and diagnosis in child psychopathology*. New York: Guilford Press.

Rutter, M., Tizard, J., & Whitmore, K. (Eds.) (1970). *Education, health and behaviour*. London: Longman. (Reprinted 1981, Melbourne: Krieger).

Rutter, M., Tizard, J.R., Yule, W., Graham, P., & Whitmore, K. (1976). Research report: Isle of Wight studies 1964-1974. *Psychological Medicine, 6*, 313-332.

Sakoda, J.M., Cohen, B.H., & Beall, G. (1954). Test of significance for a series of statistical tests. *Psychological Bulletin, 51*, 172-175.

SAS Institute (1989). *SAS User's Guide, Release 6.07.02*. Cary, NC: SAS Institute.

Sameroff, A.J., Seifer, R., Barocas, R., Zax, M., & Greenspan, S. (1987). IQ scores of 4-year-old children: Social-environmental risk factors. *Pediatrics, 79*, 343-350.

Saylor, C.F., & Brandt, B.J. (1986). The Minnesota Child Development Inventory: A valid maternal-report form for assessing development in infancy. *Developmental and Behavioral Pediatrics, 7*, 308-311.

Sherman, M., Shapiro, T., & Glassman, M. (1983). Play and language in developmentally disordered preschoolers: A new approach to classification. *Journal of the American Academy of Child Psychiatry, 22*, 511-524.

Snook, S.C., & Gorsuch, R.L. (1989). Component analysis versus common factor analysis: A Monte Carlo study. *Psychological Bulletin, 106*, 148-154.

Spiker, D., Kraemer, H.C., Constantine, N.A., & Bryant, D. (1992). Reliability and validity of behavior problem checklists as measures of stable traits in low birthweight, premature preschoolers. *Child Development*, in press.

Sroufe, L.A. (1979). The coherence of individual development. *American Psychologist, 34*, 834-841.

Sroufe, L.A. (1983). Infant-caregiver attachment and patterns of adaptation in preschool: The roots of maladaptation and competence. In M. Perlmutter (Ed.), *Minnesota Symposia in Child Psychology* (Vol. 16). Hillsdale, NJ: Erlbaum.

Sroufe, L.A. (1989). Pathways to adaptation and maladaptation: Psychopathology and developmental deviation. In D. Cicchetti (Ed.). *The emergence of a discipline: Rochester Symposium on Developmental Psychopathology, Vol. I* (pp. 13-40). Hillsdale, NJ: Erlbaum.

Sroufe, L.A., & Rutter, M. (1984). The domain of developmental psychopathology. *Child Development, 54*, 1615-1627.

Stanger, C., McConaughie, S.H., & Achenbach, T.M. (1992). Three-year course of behavioral/emotional problems in a national sample of 4- to 16-year-olds: II. Predictors of syndromes. *Journal of the American Academy of Child and Adolescent Psychiatry, 31*, 941-950.

Stanger, C., McConaughie, S.H., & Achenbach, T.M. (1993). Three-year course of behavioral/emotional probems in a national sample of 4- to 16-year-olds: III. Predictors of signs of disturbance. *Journal of Child Psychology and Psychiatry*, in press.

Stevenson, J., Richman, N., & Graham, P. (1985). Behaviour problems and language abilities at three years and behavioural deviance at eight years. *Journal of Child Psychology and Psychiatry, 26*, 215-230.

Swets-Gronert, F. (1986). *Temperament, taalcompetentie en gedragsproblemen van jonge kinderen*. Lisse: Swets & Zeitlinger.

Tanaka, J. S. (1987). "How big is big enough?": Sample size and goodness of fit in structural equation models with latent variables. *Child Development, 58*, 134-146.

Tanaka, J. S., & Huba, G. J. (1984). Hierarchical confirmatory factor analyses of psychological distress measures. *Journal of Personality and Social Psychology, 46*, 621-635.

Taylor, E. (1988). Attention deficit and conduct disorder syndromes. In M. Rutter, A.H. Tuma, & I.S. Lann (Eds.), *Assessment and diagnosis in child psychopathology* (pp. 377-407). New York: Guilford Press.

Thomas, A., & Chess, S. (1977). *Temperament and development*. New York: Brunner/Mazel.

Thompson, R.H. (1985). *Psychosocial research on pediatric hospitalization and health care: A review of the literature*. Springfield: Charles Thomas.

Tomblin, J.B., Shonrock, C.M., & Hardy, J.C. (1989). The concurrent validity of the Minnesota Child Development Inventory as a measure of young children's language development. *Journal of Speech and Hearing Disorders, 54*, 101-105.

Trad, P. V. (1988). The preschool child. New York: Wiley.

Tremblay, R. E., Desmarais-Gervais, L., Gagnon, C., & Charlebois, P. (1987). The Preschool Behaviour Checklist: Stability of its factor structure between cultures, sexes, ages and socioeconomic classes. *International Journal of Behavioral Development, 10*, 467-484.

Twijnstra, M.H., Hüsken, F.A.H., De Jonge, L.J., & Olthoff, C.L. (1992). *Eindrapportage MKD/MKT*. Nijmegen: GITP.

Ullman, D.G., & Kausch, D.F. (1979). Early identification of developmental strengths and weaknesses in preschool children. *Exceptional Children, 46*, 8-13.

Van Den Boom, D. (1988). Neonatal irritability and the development of attachment: Observation and intervention. Unpublished doctoral dissertation. State University, Leiden.

Van Den Oord, E.J.C.G. (1993). A genetic study of problem behaviors in children. Unpublished doctoral dissertation. Erasmus University, Rotterdam.

Van Den Oord, E.J.C.G., Koot, H.M., Boomsma, D.I., Verhulst, F.C., & Orlebeke, J.F. (1993). A twin-singleton comparison of problem behavior in 2-3-year olds. *Journal of Child Psychology and Psychiatry*, accepted.

Van Westerlaak, J.M., Kropman, J.A., & Collaris, J.W.M. (1975). *Beroepenklapper* [Index of occupations]. Nijmegen: Instituut voor Toegepaste Sociologie.

Van Yperen, T.A. (1990). Multi-axiale classificatie van specifieke ontwikkelingsstoornissen: Een studie over as II van het MAC. Unpublished doctoral dissertation. State University, Leiden. thesis.

Veerman, J.W., & Tates, J.G. (1987). *Handleiding voor het Scoringsschema Dossierin formatie (SDI); Episcript nr. 4.* Amsterdam: Paedologisch Instituut.

Veerman, J.W. (1990). *De ontwikkeling van kinderen na een periode van klinische jeugdhulpverlening: Een evaluatief epidemiologisch onderzoek.* Amersfoort: Academische Uitgeverij.

Velez, C.N., Johnson, J., & Cohen, P. (1989). A longitudinal analysis of selected risk factors for child psychopathology. *Journal of the American Academy of Child and Adolescent Psychiatry*, *28*, 861-864.

Verhulst, F.C. (1985). Mental health in Dutch children. Unpublished doctoral dissertation. Erasmus University, Rotterdam.

Verhulst, F.C., Achenbach, T.M., & Akkerhuis, G.W. (1989). Problems reported for clinically referred American and Dutch children. *Journal of the American Academy of Child and Adolescent Psychiatry*, *28*, 516-524.

Verhulst, F.C., Achenbach, T.M., Althaus, M., & Akkerhuis, G.W. (1988). A comparison of syndromes derived from the Child Behavior Checklist for American and Dutch girls aged 6-11 and 12-16. *Journal of Child Psychology and Psychiatry*, *29*, 879-895.

Verhulst, F.C., & Akkerhuis, G.W. (1983). Gedragsproblemen bij driejarigen. [Behavior problems in 3-year-olds.]. *Tijdschrift voor Jeugdgezondheidszorg*, *2*, 21-24.

Verhulst, F.C., & Akkerhuis, G.W. (1986). Mental health in Dutch children: III behavioral / emotional problems reported by teachers of children aged 4-12. *Acta Psychiatrica Scandinavica*, *74*, (Suppl. No. 330).

Verhulst, F.C., Akkerhuis, G.W., & Althaus. (1985a). Mental health in Dutch children: I a cross-cultural comparison. *Acta Psychiatrica Scandinavica*, *72* (Suppl. No. 323).

Verhulst, F.C., & Althaus, M. (1988). Persistence and change in behavioral/emotional problems reported by parents of children aged 4-14: An epidemiological study. *Acta Psychiatrica Scandinavica*, *77* (Suppl. No. 339).

Verhulst, F.C., Berden, G.F.M.G., & Sanders-Woudstra, J.A.R. (1985b). Mental health in Dutch children. II. The prevalence of psychiatric disorder and relationship between measures. *Acta Psychiatrica Scandinavica*, *72*, Suppl. No. 324.

Verhulst, F.C., & Koot, H.M. (1992). *Child psychiatric epidemiology: Concepts, methods, and findings.* Newbury Park, CA: Sage Publications.

Verhulst, F.C., Koot, H.M., & Berden, G.F.M.G. (1990). Four-year follow-up of an epidemiological sample. *Journal of the American Academy of Child and Adolescent Psychiatry*, *29*, 440-448.

Verhulst, F.C., Koot, J.M., Akkerhuis, G.W., & Veerman, J.W. (1990). *Praktische handleiding voor de CBCL (Child Behavior Checklist)*. Assen/Maastricht: Van Gorcum.

Verhulst, F.C., Prince, J., Vervuurt-Poot, C., & De Jong, J. (1989). Mental health in Dutch adolescents: Self-reported competencies and problems for ages 11-18. *Acta Psychiatrica Scandinavica*, *80* (Suppl. No. 356).

Verhulst, F.C., & Van Der Ende, J. (1991). Four-year follow-up of teacher-reported problem behaviours. *Psychological Medicine*, *21*, 965-977.

Weisglas-Kuperus, N., Koot, H.M., Baerts, W., Fetter, W.P.F., & Sauer, P.J.J. (1993). Behavioural problems of very low-birthweight children. *Developmental Medicine and Child Neurology*, *35*, 406-416.

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

Wenar, C. (1982). On Negativism. *Human Development*, *25*, 1-23.

Weng, J. L.-J. (1990). Aspects of covariance structure analysis with dependent observations. Unpublished doctoral dissertation, University of California, Los Angeles.

Werner, E.E., & Smith, R.S. (1982). *Vulnerable but invincible: A longitudinal study of resilient children and youth*. New York: McGraw Hill.

Windle, M. (1989). Predicting temperament-health relationships: A covariance structure latent variable analysis. *Journal of Research in Personality*, *23*, 118-144.

Windle, M. (1991). The difficult temperament in adolescence: Associations with substance use, family support, and problem behaviors. *Journal of Abnormal Psychology*, *47*, 310-315.

Windle, M. (In press). Temperament and social support in adolescence: Interrelations with depressive symptoms and delinquent behaviors. *Journal of Youth and Adolescence*.

Windle, M., Iwawaki, S., & Lerner, R.M. (Submitted). Cross-cultural comparability of temperament among Japanese and American preschool children.

Windle, M., & Lerner, R.M. (1986). Reassessing the dimensions of temperamental individuality across the life span: The Revised Dimensions of Temperament Survey (DOTS-R). *Journal of Adolescent Research*, 1, 213-230.

Windle, M., Hooker, K., Lerner, K., East, P.L., Lerner, J.V., & Lerner, R.M. (1986). Temperament, perceived competence, and depression in early- and late-adolescents. *Developmental Psychology*, *22*, 384-392.

Wolff, S. (1961). Symptomatology and outcome of pre-school children with behavior disorders attending a child guidance clinic. *Journal of Child Psychology and Psychiatry*, *2*, 269-276.

Wolkind, S.N., & DeSalis, W. (1982). Infant temperament, maternal mental stage and child behavioral problems. In R. Porter & G. M. Collins (Eds.), *Temperamental differences in infants and young children* (pp. 221-239). Ciba Foundation Symposium. London: Pitman.

World Health Organization (1988). *Draft multiaxial classification of child psychatric disorders, Axis Five: Associated abnormal psychosocial conditions.* (WHO/MNH/PRO/86.1 Rev.1). Geneva: Author.

World Health Organization (1989). *ICD-10, 1989 Draft of chapter V, Categories F00-F99: Mental and behavioral disorders (including disorders of psychological development).* (WHO/MNH/MEP/87.1 Rev.3). Geneva: Author.

Zahn-Waxler, C., Ianotti, R.J., Cummings, E.M., & Denham, S. (1990). Antecedents of problem behaviors in children of depressed mothers. *Development and Psychopathology*, 2, 271-291.

Appendix C.1
Letter to the parents of children in the community sample - Time 1

Dear parent(s) / caregiver(s),

Until now no population survey of the behavior of children ages 2-3 years has been done in the Netherlands. As researchers affiliated with the Sophia Children's Hospital we would like to ask your participation in the present population survey.

Our goal is to obtain information on several important parts of the daily life of young children, such as sleeping, eating, behavior with parents, physical complaints, and on their development and possible problems.

This information will be obtained on a large number of children ages 2-3 years in the province of Zuid-Holland. Your child has been selected completely at random for this survey. Your address has been provided by the Inoculation Register of the province of Zuid-Holland or the community of Rotterdam. It will be used only for this study.

We would like to ask your permission for a visit of one of our field-workers to your home. She will ask you a number of questions about your child. The interview will take 30-45 minutes. The field-worker will contact you (if possible) by telephone between september and december to make an appointment.

The information that we will get from you will be recorded and used completely anonymously. The information will be separated from your child's name and address. Only results about the whole group of children in the survey will be published.

We sincerely hope that you will give your cooperation in the study and that, by doing this, you want to contribute to the improvement of help to children who have problems.

Yours sincerely,

J.M. Koot, psychologist.
F.C. Verhulst, medical doctor.

Appendix C.2
Mental health agencies participating in the clinical study

1. Sophia Kinderziekenhuis Rotterdam, Afdeling Kinder- en Jeugdpsychiatrie
2. Academisch Medisch Centrum Amsterdam, Polikliniek Kinderpsychiatrie
3. Academisch Ziekenhuis Utrecht, Polikliniek Kinderpsychiatrie
4. Juliana Kinderziekenhuis, Afdeling Kinderpsychiatrie
5. Riagg Rotterdam Zuid, Afdeling Jeugd
6. Riagg Rijnmond Noord-Oost, Afdeling Kinderen en Jeugd, Sectie Capelle
7. Riagg Rijnmond Noord-Oost, Afdeling Kinderen en Jeugd, Sectie Rotterdam
8. Riagg Rijnmond Noord-West Rotterdam, Afdeling Kinderen en Jeugd
9. Medisch Kleuterdagverblijf "De Kleine Plantage" Rotterdam
10. Medisch Kleuterdagverblijf "Klavertje Vier" Rotterdam-Hoogvliet
11. Medisch Kleuterdagverblijf "Teesinkweide" Boekelo
12. Pedagogisch Consultatiebureau Rotterdam

Appendix C.3
Letter to the parents of children in the clinical sample[1]

Dear parent(s) / caregiver(s),

In order to give you and your child the best possible help, it is very important for us to be well- informed on the development and behavior of your child. To get this information we will have one or more conversations with you.

To complete our information, we would like to ask you to fill out this questionnaire. You will need 10-15 minutes to complete it.

The information that you give will be used for two purposes:

1. We know from experience that the data from the questionnaire give us very important additional information on the problems of your child. These data will help us make the advise and treatment we give to you and your child more effective.

2. The data will also be used completely anonymously in research by the Sophia Children's Hospital / Academic Hospital Rotterdam, department of Child and Adolescent Psychiatry, to which we have given our cooperation. The goal of this research is to get a better view at the problems of children who are referred to our institution.

If you have any questions about the questionnaire, or about the study you may ask your mental health worker. You are kindly requested to hand the questionnaire over to him or her once you have completed it.

Thank you for your cooperation.

.....
(Head of the department or institution).

[1] Each of the participating institutions may have made slight changes to the standard letter to the parents.

Appendix C.4
Child Behavior Checklist for Ages 2-3[1]

CHILD BEHAVIOR CHECKLIST FOR AGES 2-3

For office use only
ID #

CHILD'S NAME

SEX ☐ Boy ☐ Girl | AGE | ETHNIC GROUP OR RACE

TODAY'S DATE Mo. ____ Date ____ Yr. ____

CHILD'S BIRTHDATE Mo. ____ Date ____ Yr. ____

Please fill out this form to reflect *your* view of the child's behavior even if other people might not agree. Feel free to write additional comments beside each item and in the space provided on page 2.

PARENTS' USUAL TYPE OF WORK, even if not working now *(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: ____

THIS FORM FILLED OUT BY:

☐ Mother (name): ____

☐ Father (name): ____

☐ Other—name & relationship to child: ____

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

0 = Not True (as far as you know) 1 = Somewhat or Sometimes True 2 = Very True or Often True

0 1 2 1. Aches or pains (without medical cause)
0 1 2 2. Acts too young for age
0 1 2 3. Afraid to try new things
0 1 2 4. Avoids looking others in the eye
0 1 2 5. Can't concentrate, can't pay attention for long
0 1 2 6. Can't sit still or restless
0 1 2 7. Can't stand having things out of place
0 1 2 8. Can't stand waiting; wants everything now
0 1 2 9. Chews on things that aren't edible
0 1 2 10. Clings to adults or too dependent
0 1 2 11. Constantly seeks help
0 1 2 12. Constipated, doesn't move bowels
0 1 2 13. Cries a lot
0 1 2 14. Cruel to animals
0 1 2 15. Defiant
0 1 2 16. Demands must be met immediately
0 1 2 17. Destroys his/her own things
0 1 2 18. Destroys things belonging to his/her family or other children
0 1 2 19. Diarrhea or loose bowels when not sick
0 1 2 20. Disobedient
0 1 2 21. Disturbed by any change in routine
0 1 2 22. Doesn't want to sleep alone
0 1 2 23. Doesn't answer when people talk to him/her
0 1 2 24. Doesn't eat well (describe): ____
0 1 2 25. Doesn't get along with other children
0 1 2 26. Doesn't know how to have fun, acts like a little adult
0 1 2 27. Doesn't seem to feel guilty after misbehaving
0 1 2 28. Doesn't want to go out of home
0 1 2 29. Easily frustrated
0 1 2 30. Easily jealous
0 1 2 31. Eats or drinks things that are not food—**don't** include sweets (describe): ____
0 1 2 32. Fears certain animals, situations, or places (describe): ____
0 1 2 33. Feelings are easily hurt
0 1 2 34. Gets hurt a lot, accident-prone
0 1 2 35. Gets in many fights
0 1 2 36. Gets into everything
0 1 2 37. Gets too upset when separated from parents
0 1 2 38. Has trouble getting to sleep
0 1 2 39. Headaches (without medical cause)
0 1 2 40. Hits others
0 1 2 41. Holds his/her breath
0 1 2 42. Hurts animals or people without meaning to
0 1 2 43. Looks unhappy without good reason
0 1 2 44. Angry moods
0 1 2 45. Nausea, feels sick (without medical cause)
0 1 2 46. Nervous movements or twitching (describe): ____
0 1 2 47. Nervous, highstrung, or tense
0 1 2 48. Nightmares
0 1 2 49. Overeating
0 1 2 50. Overtired
0 1 2 51. Overweight
0 1 2 52. Painful bowel movements
0 1 2 53. Physically attacks people
0 1 2 54. Picks nose, skin, or other parts of body (describe): ____
0 1 2 55. Plays with own sex parts too much
0 1 2 56. Poorly coordinated or clumsy
0 1 2 57. Problems with eyes without medical cause (describe): ____
0 1 2 58. Punishment doesn't change his/her behavior
0 1 2 59. Quickly shifts from one activity to another
0 1 2 60. Rashes or other skin problems (without medical cause)
0 1 2 61. Refuses to eat
0 1 2 62. Refuses to play active games
0 1 2 63. Repeatedly rocks head or body
0 1 2 64. Resists going to bed at night

Please see other side

Appendix C.4 (Continued)

0 = Not True (as far as you know) 1 = Somewhat or Sometimes True 2 = Very True or Often True

0	1	2	65. Resists toilet training (describe): _____	0	1	2	82. Sudden changes in mood or feelings
			_____	0	1	2	83. Sulks a lot
0	1	2	66. Screams a lot	0	1	2	84. Talks or cries out in sleep
0	1	2	67. Seems unresponsive to affection	0	1	2	85. Temper tantrums or hot temper
0	1	2	68. Self-conscious or easily embarrassed	0	1	2	86. Too concerned with neatness or cleanliness
0	1	2	69. Selfish or won't share	0	1	2	87. Too fearful or anxious
0	1	2	70. Shows little affection toward people	0	1	2	88. Uncooperative
0	1	2	71. Shows little interest in things around him/her	0	1	2	89. Underactive, slow moving, or lacks energy
0	1	2	72. Shows too little fear of getting hurt	0	1	2	90. Unhappy, sad, or depressed
0	1	2	73. Shy or timid	0	1	2	91. Unusually loud
0	1	2	74. Sleeps less than most children during day and/or night (describe): _____	0	1	2	92. Upset by new people or situations (describe): _____
			_____				_____
0	1	2	75. Smears or plays with bowel movements	0	1	2	93. Vomiting, throwing up (without medical cause)
0	1	2	76. Speech problem (describe): _____	0	1	2	94. Wakes up often at night
			_____	0	1	2	95. Wanders away from home
0	1	2	77. Stares into space or seems preoccupied	0	1	2	96. Wants a lot of attention
0	1	2	78. Stomachaches or cramps (without medical cause)	0	1	2	97. Whining
				0	1	2	98. Withdrawn, doesn't get involved with others
0	1	2	79. Stores up things he/she doesn't need (describe): _____	0	1	2	99. Worrying
			_____				100. Please write in any problems your child has that were not listed above.
0	1	2	80. Strange behavior (describe): _____	0	1	2	_____
			_____	0	1	2	_____
0	1	2	81. Stubborn, sullen, or irritable	0	1	2	_____

PLEASE BE SURE YOU HAVE ANSWERED ALL ITEMS. UNDERLINE ANY YOU ARE CONCERNED ABOUT.

Does your child have any illness, physical disability, or mental handicap? ☐ No ☐ Yes—Please describe

What concerns you most about your child?

Please describe the best things about your child:

[1] Reproduced by permission of T.M. Achenbach.

Appendix D.1
Intercorrelations of CBCL/2-3 factors obtained from exploratory factor analysis

Clinical sample	1.	2.	3.	4.	5.	6.	7.
1. Oppositional	1.000						
2. Withdrawn/Depressed	.049	1.000					
3. Aggressive	.376	.211	1.000				
4. Anxious	.136	.300	.180	1.000			
5. Overactive	.260	-.007	.296	-.067	1.000		
6. Sleep Problems	.236	.010	.171	.199	.121	1.000	
7. Somatic Problems	.252	.181	.190	.236	-.081	.123	1.000

Community sample	1.	2.	3.	4.	5.	6.	7.
1. Oppositional	1.000						
2. Withdrawn/Depressed	.292	1.000					
3. Aggressive	.286	.235	1.000				
4. Anxious	.282	.352	.008	1.000			
5. Overactive	.404	.208	.375	.177	1.000		
6. Sleep Problems	.209	.091	.059	.210	.183	1.000	
7. Somatic Problems	.063	-.038	.039	.047	.093	.134	1.000

Twin sample	1.	2.	3.	4.	5.	6.	7.
1. Oppositional	1.000						
2. Withdrawn/Depressed	.311	1.000					
3. Aggressive	.555	.165	1.000				
4. Anxious	.414	.409	.135	1.000			
5. Overactive	.384	.023	.361	.160	1.000		
6. Sleep Problems	.459	.274	.337	.270	.195	1.000	
7. Somatic Problems	.297	.295	.120	.264	.097	.394	1.000

Appendix D.2

CBCL/2-3 items loading ≥.30 on factors recurring in exploratory factor analyses in the clinical, community and twin sample

Factor / Items	Clinical sample	Community sample	Twin sample
I. Oppositional			
8. Can't wait	.57	.33 .39 [5]	.68
13. Cries much	.46	.32	.43
15. Defiant	.45	.35	.32 .31 [3]
16. Demands must be met	.63	.37 .35 [5]	.73
20. Disobedient	.52	.-	.38 .30 [3]
29. Easily frustrated	.47	.-	.49
30. Easily jealous	.34 .32 [3]	.-	.48
33. Feelings are easily hurt	.-	.36	.52
36. Gets into everything	.-	.36	.42
44. Angry moods	.69	.53	.75
66. Screams	.57	.31	.42
69. Selfish	.32 .37 [3]	.35	.34
81. Stubborn	.72	.67	.73
82. Moody	.54	.51	.62
83. Sulks	.48	.59	.69
85. Temper tantrums	.76	.57	.76
88. Uncooperative	.55	.-	.30
91. Too loud	.42	.35	.36
96. Wants attention	.43	.37 .41 [5]	.42
97. Whining	.57	.53	.54
Eigenvalue	14.63	11.05	15.49
II. Withdrawn/Depressed			
2. Acts too young	.63	.41	.48
23. Doesn't answer	.50	.32	.-
26. No fun	.-	.30	.34
43. Looks unhappy	.-	.58	.44
56. Clumsy	.30 .39 [5]	.35	.42 .35 [5]
67. Unresponsive to affection	.40	.37	.46
70. Little affection	.38 .30 [3]	.47	.63
71. Little interest	.51	.47	.59
76. Speech problem	.55	.36	.-
77. Stares blankly	.41	.48	.-
80. Strange behavior	.-	.41	.36
89. Underactive	.-	.35	.42
90. Sad	.-	.32	.48
98. Withdrawn	.50	.54 .30 [4]	.-
Eigenvalue	2.24	3.70	1.91
I. Aggressive			
4. Cruel to animals	.47	.54	.41
17. Destroys own things	.32	.54	.72
18. Destroys other's things	.36	.65	.73
35. Fights	.64	.44	.43
40. Hits	.54 .31 [1]	.47	.51
2. Hurts accidentally	.42	.50	.46
3. Attacks people	.58	.41	.45
igenvalue	5.40	2.59	3.77

(continued)

Appendix D.2 (Continued)

Factor / Items	Clinical sample		Community sample	Twin sample
IV. Anxious				
3. Afraid to try new things	.57		.54	.56
4. Avoids eye contact	.-		.40	.61
10. Clings to adults	.52	.32 [5]	.43	.53
21. Disturbed by change	.37		.32	.38
32. Fears	.47		.-	.38
37. Upset by separation	.41		.42	.45
68. Self-conscious	.-		.44	.49
73. Shy	.48		.61	.74
87. Too fearful or anxious	.61		.45	.56
92. Upset by new	.62		.53	.72
Eigenvalue	3.38		2.28	1.96
V. Overactive				
5. Can't concentrate	.64		.67	.66
6. Can't sit still	.44	.40 [1]	.65	.49
11. Constantly seeks help	.39		.41	.35
59. Quickly shifts activity	.52	.34 [1]	.61	.59
62. Refuses active games	.44		.52	.34
Eigenvalue	1.65		1.64	1.29
VI. Sleep Problems				
22. Doesn't want to sleep alone	.59		.59	.-
38. Can't sleep	.64		.62	.62
48. Nightmares	.47		.51	.61
64. Resists going to bed	.51		.38	.48
74. Sleeps little	.54		.43	.48
84. Talks or cries in sleep	.37		.40	.48
94. Wakes often	.56		.65	.73
Eigenvalue	1.98		1.54	1.46
VII. Somatic Problems				
1. Aches	.40		.-	.50
12. Constipated	.40		.39	.-
45. Nausea	.42		.46	.54
52. Painful bowel movements	.47		.37	.34
78. Stomachaches	.43		.30	.60
93. Vomiting	.41		.37	.36
Eigenvalue	1.43		1.44	1.07

Note: Loadings are standardized regression coefficients obtained from promax rotations. Cross-loadings are given, followed by the number of the factor [in brackets] on which the cross-loading occurred.

Appendix D.3
Intercorrelations of observed CBCL/2-3 scale scores

Clinical sample	1.	2.	3.	4.	5.	6.	7.
1. Oppositional	1.000						
2. Withdrawn/Depressed	.363	1.000					
3. Aggressive	.643	.275	1.000				
4. Anxious	.360	.537	.154	1.000			
5. Overactive	.554	.308	.545	.213	1.000		
6. Sleep Problems	.365	.119	.200	.268	.193	1.000	
7. Somatic Problems	.200	.242	.150	.188	.132	.222	1.000

Community sample	1.	2.	3.	4.	5.	6.	7.
1. Oppositional	1.000						
2. Withdrawn/Depressed	.357	1.000					
3. Aggressive	.490	.276	1.000				
4. Anxious	.348	.426	.124	1.000			
5. Overactive	.530	.305	.397	.193	1.000		
6. Sleep Problems	.220	.195	.143	.243	.173	1.000	
7. Somatic Problems	.206	.133	.103	.142	.094	.234	1.000

Twin sample	1.	2.	3.	4.	5.	6.	7.
1. Oppositional	1.000						
2. Withdrawn/Depressed	.475	1.000					
3. Aggressive	.653	.401	1.000				
4. Anxious	.486	.484	.251	1.000			
5. Overactive	.579	.374	.545	.306	1.000		
6. Sleep Problems	.392	.249	.291	.294	.271	1.000	
7. Somatic Problems	.241	.209	.160	.195	.144	.219	1.000

Appendix D.4

Factor loadings for CBCL/2-3 scales on a second order factor obtained from exploratory factor analyses in the clinical, community and twin sample for the one-factor solution

Scales	Clinical sample	Community sample	Twin sample
Oppositional	.849	.792	.874
Withdrawn/Depressed	.523	.544	.609
Aggressive	.662	.541	.690
Anxious	.474	.459	.551
Overactive	.638	.604	.654
Sleep Problems	.381	.341	.452
Somatic Problems	.296	.258	.294

Note. Factor loadings are standardized regression coefficients obtained from promax rotations.

Appendix D.5

Factor loadings for CBCL/2-3 scales on second order factors obtained from exploratory factor analyses in the clinical, community, and twin sample

	Clinical sample		Community sample		Twin sample	
Factor	Factor I	Factor II	Factor I	Factor II	Factor I	Factor II
Difficult	.759	.187	.699	.183	.669	.291
Withdrawn/Depressed	.163	.550	.203	.461	.248	.471
Aggressive	.817	-.06	.652	-.05	.855	-.07
Anxious	-.07	.900	-.10	.801	-.06	.827
Overactive	.650	.061	.638	.035	.628	.092
Sleep Problems	.235	.219	.101	.313	.246	.270
Somatic Problems	.131	.237	.093	.213	.104	.240

Note. Factor loadings are standardized regression coefficients obtained from promax rotations.

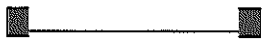
Appendix E.1
Mean CBCL/2-3 scale scores for matched referred and nonreferred children

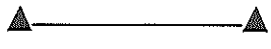
	Referred		Nonreferred		*SE* of the mean		*SE* of measurement		Alpha
Scale	Mean	*SD*	Mean	*SD*	Ref	Nonref	Ref	Nonref	
Oppositional	17.7	8.5	10.7	6.1	.47	.34	2.96	2.11	.91
Withdrawn/Depressed	4.1	3.5	1.1	1.7	.19	.10	2.19	1.09	.77
Aggressive	6.2	4.4	3.4	2.7	.25	.15	1.71	1.05	.85
Anxious	5.2	4.1	3.3	2.9	.23	.16	1.68	1.21	.80
Overactive	5.6	2.8	3.2	2.5	.16	.14	1.14	.99	.81
Sleep Problems	4.1	3.8	2.1	2.6	.21	.14	1.84	1.28	.81
Somatic Problems	0.5	0.9	0.3	0.6	.05	.03	0.47	0.32	.42
Internalizing	9.4	6.6	4.5	4.0	.37	.22	2.89	1.75	.85
Externalizing	29.5	13.9	17.3	9.4	.77	.52	4.39	2.98	.94
Total Problems	60.6	26.5	34.0	17.2	1.47	.95	9.54	6.19	.95

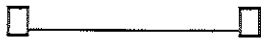
Note. $N = 326$ each in referred and nonreferred samples matched by sex, age, and SES.
SE of measurement = $SD\sqrt{(1\text{-reliability})}$, computed from test-retest reliabilities reported in Chapter 4.

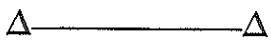
Appendix E.2
Mean scores on CBCL/2-3 scales and Total Problems according to referral status, sex, and age

Legenda of the Figures:

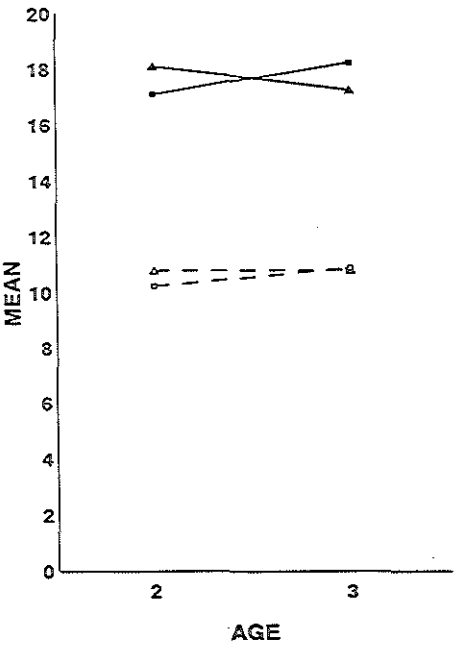

Referred Boys


Referred Girls

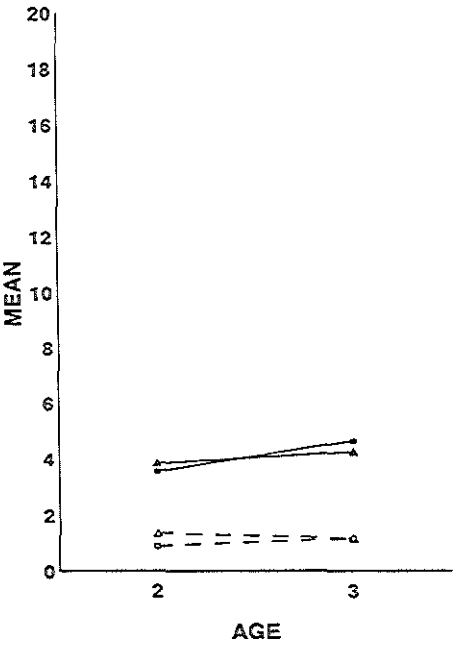

Nonreferred Boys


Nonreferred Girls

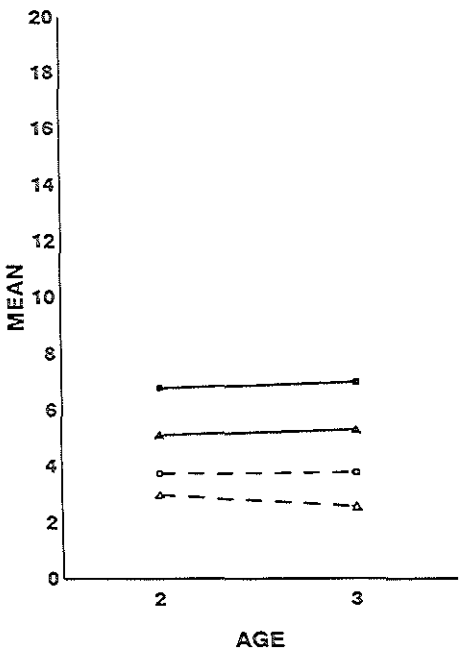
OPPOSITIONAL



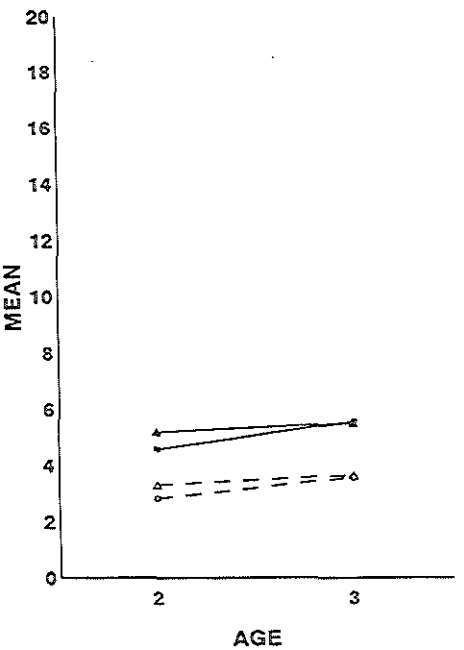
WITHDRAWN/DEPRESSED

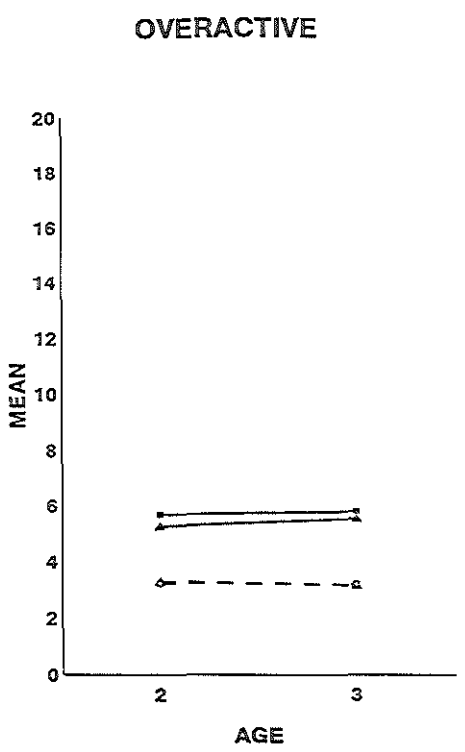


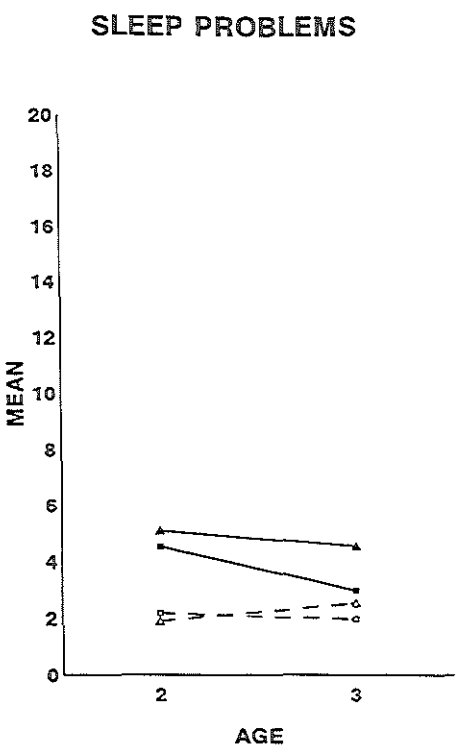
AGGRESSIVE

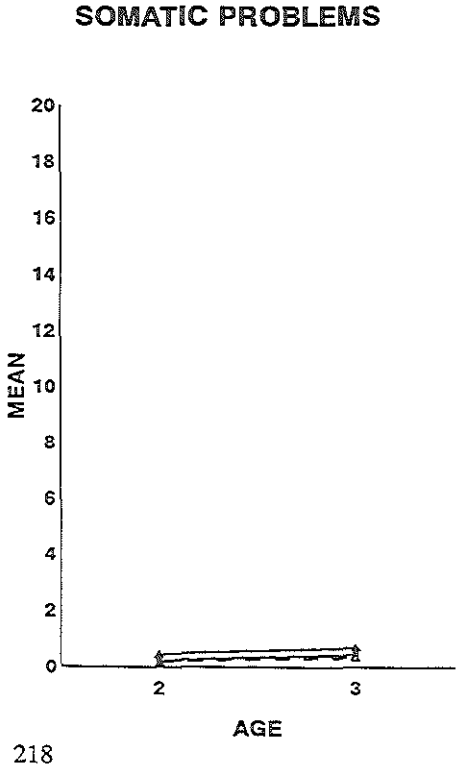


ANXIOUS

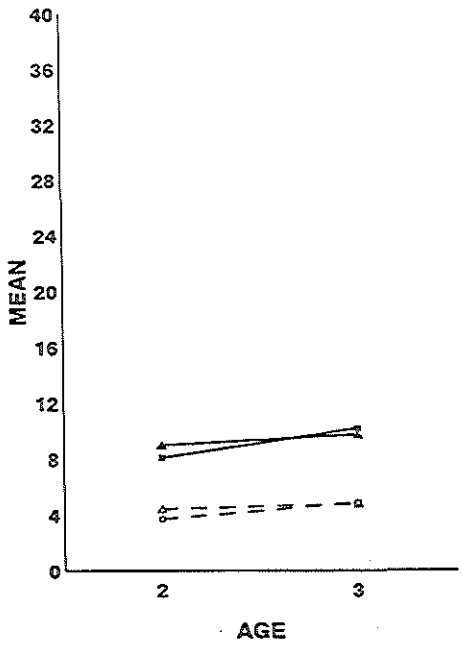



OVERACTIVE
MEAN
20
18
16
14
12
10
8
6
4
2
0
2
3
AGE

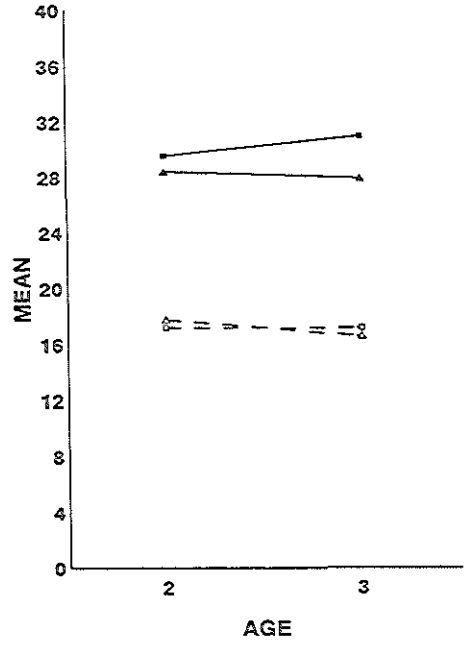

SLEEP PROBLEMS
MEAN
20
18
16
14
12
10
8
6
4
2
0
2
3
AGE


SOMATIC PROBLEMS
MEAN
20
18
16
14
12
10
8
6
4
2
0
2
3
AGE

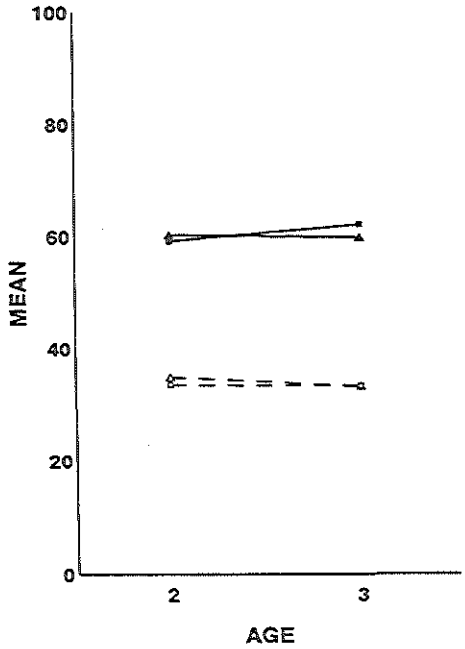
INTERNALIZING



EXTERNALIZING



TOTAL PROBLEMS



Appendix E.3
Percentage of CBCL/2-3 item scores 1 and 2, mean item scores, and standard error of the mean for children grouped [a] by referral status and sex

Item		Group	1	2	1+2	Mean	SE[b]
1.	Aches	RB	10.6	2.5	13.1	.16	.03
		RG	20.3	6.3	26.6	.33	.05
		NB	11.1	2.5	13.6	.16	.03
		NG	18.8	0.0	18.8	.19	.04
2.	Acts too young	RB	16.2	20.7	36.9	.58	.06
		RG	8.6	22.7	31.3	.54	.07
		NB	8.1	3.0	11.1	.14	.03
		NG	5.5	0.8	6.3	.07	.03
3.	Afraid to try new things	RB	24.7	12.6	37.3	.50	.05
		RG	28.9	10.9	39.8	.51	.06
		NB	27.8	7.1	34.9	.42	.04
		NG	27.3	5.5	32.8	.38	.05
4.	Avoids eye contact	RB	27.8	11.1	38.9	.50	.05
		RG	25.0	6.3	31.3	.38	.05
		NB	23.2	1.5	24.7	.26	.03
		NG	21.9	1.6	23.5	.25	.04
5.	Can't concentrate	RB	30.8	43.9	74.7	1.19	.06
		RG	29.7	44.5	74.2	1.19	.07
		NB	35.9	11.6	47.5	.59	.05
		NG	33.6	10.2	43.8	.54	.06
6.	Can't sit still	RB	26.8	56.6	83.4	1.40	.05
		RG	25.8	46.9	72.7	1.20	.07
		NB	35.9	24.7	60.6	.85	.06
		NG	36.7	24.2	60.9	.85	.07
7.	Can't stand having things moved	RB	19.2	15.2	34.4	.50	.05
		RG	25.0	10.9	35.9	.47	.06
		NB	15.7	11.6	27.3	.39	.05
		NG	14.1	4.7	18.8	.23	.05
8.	Can't wait	RB	28.3	59.6	87.9	1.48	.05
		RG	34.4	54.7	89.1	1.44	.06
		NB	43.4	25.3	68.7	.94	.05
		NG	44.5	28.1	72.6	1.01	.07
9.	Chews nonfood	RB	23.7	13.1	36.8	.50	.05
		RG	25.8	18.0	43.8	.62	.07
		NB	19.7	6.6	26.3	.33	.04
		NG	26.6	3.9	30.5	.34	.05
10.	Clings to adults	RB	30.3	22.2	52.5	.75	.06
		RG	37.5	22.7	60.2	.83	.07
		NB	35.4	6.6	42.0	.49	.04
		NG	35.9	7.8	43.7	.52	.06

Item	Group	1	2	1+2	Mean	SE[b]
11. Constantly seeks help	RB	38.4	23.7	62.1	.86	.06
	RG	33.6	26.6	60.2	.87	.07
	NB	36.4	6.1	42.5	.49	.04
	NG	41.4	4.7	46.1	.51	.05
12. Constipated	RB	4.5	4.0	8.5	.13	.03
	RG	4.7	11.7	16.4	.28	.06
	NB	7.1	2.5	9.6	.12	.03
	NG	8.6	3.1	11.7	.15	.04
13. Cries much	RB	33.3	15.7	49.0	.65	.05
	RG	26.6	28.9	55.5	.84	.08
	NB	23.7	2.5	26.2	.29	.04
	NG	23.4	3.9	27.3	.31	.05
14. Cruel to animals	RB	18.2	7.6	25.8	.33	.04
	RG	7.8	3.9	11.7	.16	.04
	NB	10.1	0.5	10.6	.11	.02
	NG	8.6	0.0	8.6	.09	.03
15. Defiant	RB	31.8	48.5	80.3	1.29	.06
	RG	32.0	44.5	76.5	1.21	.07
	NB	49.0	31.8	80.8	1.13	.05
	NG	53.1	25.0	78.1	1.03	.06
16. Demands must be met	RB	31.8	58.6	90.4	1.49	.05
	RG	29.7	55.5	85.2	1.41	.07
	NB	55.1	22.2	77.3	1.00	.05
	NG	50.8	24.2	75.0	.99	.06
17. Destroys own things	RB	29.3	24.2	53.5	.78	.06
	RG	27.3	15.6	42.9	.59	.07
	NB	28.8	6.1	34.9	.41	.04
	NG	24.2	0.0	24.2	.24	.04
18. Destroys others' things	RB	29.3	22.2	51.5	.74	.06
	RG	19.5	14.8	34.3	.49	.07
	NB	24.7	3.0	27.7	.31	.04
	NG	15.6	0.0	15.6	.16	.03
19. Diarrhea	RB	15.2	13.1	28.3	.42	.05
	RG	15.6	9.4	25.0	.34	.06
	NB	14.6	6.1	20.7	.27	.04
	NG	9.4	0.8	10.2	.11	.03
20. Disobedient	RB	39.4	51.0	90.4	1.41	.05
	RG	35.2	44.5	79.7	1.24	.07
	NB	71.7	12.1	83.8	.96	.04
	NG	67.2	6.3	73.5	.80	.05

Item	Group	1	2	1+2	Mean	SE[b]
21. Disturbed by change	RB	34.3	27.8	62.1	.90	.06
	RG	35.2	27.3	62.5	.90	.07
	NB	29.8	8.1	37.9	.46	.05
	NG	36.7	4.7	41.4	.46	.05
22. Doesn't want to sleep alone	RB	18.2	14.1	32.3	.47	.05
	RG	14.8	22.7	37.5	.60	.07
	NB	12.1	12.6	24.7	.37	.05
	NG	12.5	10.9	23.4	.34	.06
23. Doesn't answer	RB	41.9	25.3	67.2	.92	.05
	RG	46.1	17.2	63.3	.81	.06
	NB	36.4	4.0	40.4	.44	.04
	NG	42.2	3.1	45.3	.48	.05
24. Doesn't eat well	RB	29.8	17.7	47.5	.65	.05
	RG	28.1	26.6	54.7	.81	.07
	NB	34.3	17.7	52.0	.70	.05
	NG	39.8	17.2	57.0	.74	.07
25. Doesn't get along with other kids	RB	35.4	13.6	49.0	.63	.05
	RG	25.8	9.4	35.2	.45	.06
	NB	12.6	1.0	13.6	.15	.03
	NG	12.5	0.0	12.5	.13	.03
26. No fun	RB	14.1	2.0	16.1	.18	.03
	RG	10.2	7.8	18.0	.26	.05
	NB	2.5	0.5	3.0	.04	.02
	NG	3.1	0.8	3.9	.05	.02
27. Lacks guilt	RB	34.3	32.3	66.6	.99	.06
	RG	26.6	32.0	58.6	.91	.08
	NB	36.4	8.6	45.0	.54	.05
	NG	30.5	4.7	35.2	.40	.05
28. Doesn't want to go out	RB	9.1	1.5	10.6	.12	.03
	RG	7.0	4.7	11.7	.16	.04
	NB	2.5	2.0	4.5	.07	.02
	NG	0.8	1.6	2.4	.04	.02
29. Easily frustrated	RB	30.8	43.4	74.2	1.18	.06
	RG	27.3	38.3	65.6	1.04	.08
	NB	39.4	7.1	46.5	.54	.05
	NG	39.1	6.3	45.4	.52	.05
30. Easily jealous	RB	35.9	27.8	63.7	.91	.06
	RG	32.0	34.4	66.4	1.01	.07
	NB	38.9	7.6	46.5	.54	.05
	NG	41.4	12.5	53.9	.66	.06

Item	Group	1	2	1+2	Mean	SE[b]
31. Eats or drinks nonfood	RB	9.6	5.6	15.2	.21	.04
	RG	8.6	8.6	17.2	.26	.05
	NB	3.5	0.5	4.0	.05	.02
	NG	7.0	1.6	8.6	.10	.03
32. Fears	RB	27.3	22.7	50.0	.73	.06
	RG	28.9	22.7	51.6	.74	.07
	NB	39.9	14.1	54.0	.68	.05
	NG	31.3	21.9	53.2	.75	.07
33. Feelings easily hurt	RB	26.8	25.3	52.1	.77	.06
	RG	30.5	28.1	58.6	.87	.07
	NB	39.4	11.6	51.0	.63	.05
	NG	48.4	14.1	62.5	.77	.06
34. Accident-prone	RB	23.7	23.2	46.9	.70	.06
	RG	25.0	24.2	49.2	.73	.07
	NB	37.9	12.1	50.0	.62	.05
	NG	21.9	14.1	36.0	.50	.07
35. Fights	RB	24.7	18.2	42.9	.61	.06
	RG	24.2	8.6	32.8	.41	.06
	NB	28.3	5.6	33.9	.39	.04
	NG	18.8	1.6	20.4	.22	.04
36. Gets into everything	RB	27.8	29.8	57.6	.87	.06
	RG	22.7	38.3	61.0	.99	.08
	NB	41.4	21.2	62.6	.84	.05
	NG	43.0	20.3	63.3	.84	.07
37. Upset by separation	RB	29.8	18.7	48.5	.67	.06
	RG	33.6	18.8	52.4	.71	.07
	NB	20.7	3.5	24.2	.28	.04
	NG	20.3	5.5	25.8	.31	.05
38. Can't sleep	RB	26.8	19.2	46.0	.65	.06
	RG	14.8	34.4	49.2	.84	.08
	NB	16.7	6.1	22.8	.29	.04
	NG	14.1	6.3	20.4	.27	.05
39. Headaches	RB	2.5	1.5	4.0	.06	.02
	RG	3.1	1.6	4.7	.06	.03
	NB	2.5	0.0	2.5	.03	.01
	NG	3.1	0.0	3.1	.03	.02
40. Hits	RB	40.9	29.8	70.7	1.01	.06
	RG	35.9	19.5	55.4	.75	.07
	NB	48.5	8.1	56.6	.65	.05
	NG	43.0	1.6	44.6	.46	.05

Item	Group	1	2	1+2	Mean	SE[b]
41. Holds breath	RB	5.1	1.0	6.1	.07	.02
	RG	4.7	6.3	11.0	.17	.05
	NB	3.0	3.5	6.5	.10	.03
	NG	7.0	1.6	8.6	.10	.03
42. Hurts accidentally	RB	31.8	17.2	49.0	.66	.05
	RG	25.0	12.5	37.5	.50	.06
	NB	27.3	3.5	30.8	.34	.04
	NG	25.8	2.3	28.1	.31	.05
43. Looks unhappy	RB	17.7	7.6	25.3	.33	.04
	RG	19.5	10.2	29.7	.40	.06
	NB	5.1	0.0	5.1	.05	.02
	NG	5.5	0.0	5.5	.06	.02
44. Angry moods	RB	31.3	42.9	74.2	1.17	.06
	RG	28.9	42.2	71.1	1.13	.07
	NB	60.1	12.1	72.2	.84	.04
	NG	57.8	7.8	65.6	.73	.05
45. Nausea	RB	6.6	1.5	8.1	.10	.02
	RG	5.5	3.9	9.4	.13	.04
	NB	3.0	0.5	3.5	.04	.02
	NG	4.7	0.0	4.7	.05	.02
46. Nervous movements	RB	8.6	12.6	21.2	.34	.05
	RG	7.8	15.6	23.4	.39	.07
	NB	3.5	1.5	5.0	.07	.02
	NG	6.3	2.3	8.6	.11	.03
47. Nervous	RB	20.7	15.7	36.4	.52	.05
	RG	27.3	11.7	39.0	.51	.06
	NB	7.6	0.0	7.6	.08	.02
	NG	12.5	1.6	14.1	.16	.04
48. Nightmares	RB	22.7	9.1	31.8	.41	.05
	RG	25.0	12.5	37.5	.50	.06
	NB	17.7	3.5	21.2	.25	.04
	NG	18.0	3.1	21.1	.24	.04
49. Overeating	RB	7.6	7.1	14.7	.22	.04
	RG	12.5	4.7	17.2	.22	.05
	NB	6.1	2.0	8.1	.10	.03
	NG	4.7	2.3	7.0	.09	.03
50. Overtired	RB	26.3	10.1	36.4	.47	.05
	RG	23.4	16.4	39.8	.56	.07
	NB	15.7	3.5	19.2	.23	.04
	NG	21.1	2.3	23.4	.26	.04

Item	Group	1	2	1+2	Mean	SE[b]
51. Overweight	RB	6.6	1.0	7.6	.09	.02
	RG	5.5	1.6	7.1	.09	.03
	NB	2.0	0.0	2.0	.02	.01
	NG	2.3	2.3	4.6	.07	.03
52. Painful bowel movements	RB	4.0	1.5	5.5	.07	.02
	RG	10.2	6.3	16.5	.23	.05
	NB	3.0	1.0	4.0	.05	.02
	NG	7.0	1.6	8.6	.10	.03
53. Attacks people	RB	20.2	10.6	30.8	.41	.05
	RG	14.8	6.3	21.1	.27	.05
	NB	8.6	1.5	10.1	.12	.03
	NG	5.5	0.0	5.5	.06	.02
54. Picking	RB	17.2	16.7	33.9	.51	.05
	RG	22.7	18.8	41.5	.60	.07
	NB	20.7	9.1	29.8	.39	.05
	NG	25.0	10.2	35.2	.45	.06
55. Plays with sex parts	RB	15.2	7.6	22.8	.30	.04
	RG	9.4	5.5	14.9	.20	.05
	NB	12.6	1.5	14.1	.16	.03
	NG	12.5	1.6	14.1	.16	.04
56. Clumsy	RB	15.7	18.7	34.4	.53	.06
	RG	17.2	16.4	33.6	.50	.07
	NB	10.1	3.5	13.6	.17	.03
	NG	12.5	0.8	13.3	.14	.03
57. Eye problems	RB	2.5	1.0	3.5	.05	.02
	RG	0.8	0.8	1.6	.02	.02
	NB	0.0	0.0	0.0	.00	.00
	NG	0.8	0.0	0.8	.02	.02
58. Punishment doesn't change behavior	RB	33.3	51.5	84.8	1.36	.05
	RG	32.8	38.3	71.1	1.09	.07
	NB	38.4	17.2	55.6	.73	.05
	NG	37.5	10.2	47.7	.58	.06
59. Quickly shifts activity	RB	19.2	65.7	84.9	1.51	.05
	RG	27.3	57.0	84.3	1.41	.07
	NB	42.9	27.8	70.7	.99	.05
	NG	46.9	25.8	72.7	.98	.07
60. Rashes	RB	8.1	9.1	17.2	.26	.04
	RG	14.1	8.6	22.7	.31	.06
	NB	8.1	3.5	11.6	.15	.03
	NG	9.4	7.0	16.4	.23	.05

Item	Group	1	2	1+2	Mean	SE[b]
61. Refuses to eat	RB	40.4	12.6	53.0	.66	.05
	RG	35.2	21.1	56.3	.77	.07
	NB	38.4	8.1	46.5	.55	.05
	NG	35.9	7.8	43.7	.52	.06
62. Refuses active games	RB	28.8	25.3	54.1	.79	.06
	RG	36.7	16.4	53.1	.70	.07
	NB	20.2	4.5	24.7	.29	.04
	NG	25.0	4.7	29.7	.34	.05
63. Rocks head or body	RB	11.6	9.1	20.7	.30	.05
	RG	11.7	11.7	23.4	.35	.06
	NB	4.5	3.0	7.5	.11	.03
	NG	2.3	2.3	4.6	.07	.03
64. Resists going to bed	RB	26.8	10.1	36.9	.47	.05
	RG	19.5	18.8	38.3	.57	.07
	NB	12.1	2.0	14.1	.16	.03
	NG	16.4	5.5	21.9	.27	.05
65. Resists toilet training	RB	17.7	27.3	45.0	.72	.06
	RG	18.8	21.1	39.9	.61	.07
	NB	13.1	14.6	27.7	.42	.05
	NG	12.5	7.0	19.5	.27	.05
66. Screams	RB	35.4	39.4	74.8	1.14	.06
	RG	26.6	38.3	64.9	1.03	.08
	NB	35.9	10.6	46.5	.57	.05
	NG	35.2	7.8	43.0	.51	.06
67. Unresponsive to affection	RB	20.7	7.6	28.3	.36	.04
	RG	11.7	7.8	19.5	.27	.05
	NB	5.6	1.0	6.6	.08	.02
	NG	7.0	2.3	9.3	.12	.03
68. Self-conscious	RB	14.1	8.1	22.2	.30	.04
	RG	14.1	5.5	19.6	.25	.05
	NB	24.7	1.5	26.2	.28	.03
	NG	29.7	2.3	32.0	.34	.05
69. Selfish	RB	26.3	13.6	39.9	.54	.05
	RG	25.0	10.2	35.2	.45	.06
	NB	35.9	1.0	36.9	.38	.04
	NG	29.7	1.6	31.3	.33	.05
70. Little affection	RB	28.8	8.1	36.9	.45	.05
	RG	14.1	6.3	20.4	.27	.05
	NB	9.6	0.5	10.1	.11	.02
	NG	6.3	0.8	7.1	.08	.03

Item	Group	1	2	1+2	Mean	SE[b]
71. Little interest	RB	13.1	6.1	19.2	.25	.04
	RG	15.6	7.8	23.4	.31	.05
	NB	4.0	2.0	6.0	.08	.02
	NG	5.5	0.8	6.3	.07	.03
72. Little fear	RB	22.2	26.8	49.0	.76	.06
	RG	25.0	20.3	45.3	.66	.07
	NB	26.8	11.6	38.4	.50	.05
	NG	24.2	4.7	28.9	.34	.05
73. Shy	RB	24.2	10.6	34.8	.46	.05
	RG	39.8	7.8	47.6	.56	.06
	NB	35.9	7.6	43.5	.51	.05
	NG	49.2	7.8	57.0	.65	.06
74. Sleeps little	RB	12.1	20.2	32.3	.53	.06
	RG	21.1	27.3	48.4	.76	.08
	NB	8.6	4.5	13.1	.18	.04
	NG	7.8	8.6	16.4	.25	.05
75. Smears bowel movements	RB	8.1	4.5	12.6	.17	.03
	RG	7.8	3.1	10.9	.14	.04
	NB	0.5	0.0	0.5	.01	.01
	NG	0.0	0.0	0.0	.00	.00
76. Speech problem	RB	13.1	39.9	53.0	.93	.07
	RG	11.7	24.2	35.9	.60	.08
	NB	11.1	8.1	19.2	.27	.04
	NG	7.8	6.3	14.1	.20	.05
77. Stares blankly	RB	23.7	11.6	35.3	.47	.05
	RG	21.1	8.6	29.7	.38	.06
	NB	9.1	1.5	10.6	.12	.03
	NG	13.3	0.0	13.3	.13	.03
78. Stomachaches	RB	7.1	1.5	8.6	.10	.03
	RG	6.3	3.9	10.2	.14	.04
	NB	4.5	1.0	5.5	.07	.02
	NG	3.1	0.8	3.9	.05	.02
79. Stores up unneeded things	RB	8.1	4.5	12.6	.17	.03
	RG	7.8	4.7	12.5	.17	.04
	NB	11.1	8.6	19.7	.28	.04
	NG	20.3	4.7	25.0	.30	.05
80. Strange behavior	RB	15.2	12.6	27.8	.40	.05
	RG	12.5	20.3	32.8	.53	.07
	NB	2.5	0.5	3.0	.04	.02
	NG	3.1	0.8	3.9	.05	.02

Item	Group	1	2	1+2	Mean	SE[b]
81. Stubborn	RB	29.8	45.5	75.3	1.21	.06
	RG	35.2	37.5	72.7	1.10	.07
	NB	51.0	10.6	61.6	.72	.05
	NG	51.6	9.4	61.0	.70	.06
82. Moody	RB	29.8	33.8	63.6	.98	.06
	RG	27.3	36.7	64.0	1.01	.08
	NB	28.8	5.6	34.4	.40	.04
	NG	32.8	3.9	36.7	.41	.05
83. Sulks	RB	33.8	16.7	50.5	.67	.05
	RG	35.2	22.7	57.9	.81	.07
	NB	41.9	5.1	47.0	.52	.04
	NG	47.7	4.7	52.4	.57	.05
84. Talks or cries in sleep	RB	23.7	8.1	31.8	.40	.05
	RG	28.9	11.7	40.6	.52	.06
	NB	23.2	4.0	27.2	.31	.04
	NG	25.0	5.5	30.5	.36	.05
85. Temper tantrums	RB	31.3	52.5	83.8	1.36	.05
	RG	29.7	41.4	71.1	1.13	.07
	NB	42.4	11.6	54.0	.66	.05
	NG	43.8	9.4	53.2	.63	.06
86. Too concerned with neat or clean	RB	17.2	6.1	23.3	.29	.04
	RG	23.4	7.8	31.2	.39	.06
	NB	15.2	3.5	18.7	.22	.04
	NG	15.6	3.1	18.7	.22	.04
87. Too fearful or anxious	RB	19.7	12.6	32.3	.45	.05
	RG	29.7	13.3	43.0	.56	.06
	NB	17.7	2.5	20.2	.23	.03
	NG	21.9	2.3	24.2	.27	.04
88. Uncooperative	RB	48.0	26.3	74.3	1.01	.05
	RG	47.7	21.1	68.8	.90	.06
	NB	40.9	2.0	42.9	.45	.04
	NG	45.3	2.3	47.6	.50	.05
89. Underactive	RB	8.1	6.1	14.2	.20	.04
	RG	11.7	7.0	18.7	.26	.05
	NB	2.5	1.0	3.5	.05	.02
	NG	3.9	0.0	3.9	.04	.02
90. Sad	RB	14.6	4.0	18.6	.23	.04
	RG	22.7	4.7	27.4	.32	.05
	NB	6.6	0.0	6.6	.07	.02
	NG	6.3	0.0	6.3	.06	.02

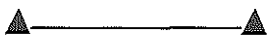
Item	Group	1	2	1+2	Mean	SE[b]
91. Too loud	RB	31.3	31.3	62.6	.94	.06
	RG	30.5	24.2	54.7	.79	.07
	NB	34.8	5.6	40.4	.46	.04
	NG	31.3	4.7	36.0	.41	.05
92. Upset by new people or situations	RB	26.3	16.7	43.0	.60	.05
	RG	25.8	21.1	46.9	.68	.07
	NB	25.8	3.0	28.8	.32	.04
	NG	24.2	3.9	28.1	.32	.05
93. Vomiting	RB	6.6	3.5	10.1	.14	.03
	RG	7.0	5.5	12.5	.18	.05
	NB	4.5	0.5	5.0	.06	.02
	NG	1.6	0.8	2.4	.03	.02
94. Wakes often	RB	22.7	25.8	48.5	.74	.06
	RG	18.8	39.8	58.6	.98	.08
	NB	19.7	14.1	33.8	.48	.05
	NG	32.8	9.4	42.2	.52	.06
95. Wanders away from home	RB	10.6	4.5	15.1	.20	.04
	RG	5.5	3.1	8.6	.12	.04
	NB	9.1	3.0	12.1	.15	.03
	NG	3.9	1.6	5.5	.07	.03
96. Wants attention	RB	18.7	64.6	83.3	1.48	.05
	RG	21.1	68.8	89.9	1.59	.06
	NB	47.0	21.2	68.2	.89	.05
	NG	46.1	23.4	69.5	.93	.07
97. Whining	RB	32.3	26.8	59.1	.86	.06
	RG	36.7	27.3	64.0	.91	.07
	NB	31.8	2.5	34.3	.37	.04
	NG	38.3	5.5	43.8	.49	.05
98. Withdrawn	RB	21.7	11.6	33.3	.45	.05
	RG	18.8	1.6	20.4	.22	.04
	NB	6.1	2.0	8.1	.10	.03
	NG	7.0	0.0	7.0	.07	.02
99. Worrying	RB	10.1	4.5	14.6	.19	.04
	RG	9.4	3.1	12.5	.16	.04
	NB	6.6	0.0	6.6	.07	.02
	NG	9.4	0.0	9.4	.09	.03
100. Other problems	RB	8.1	11.6	19.7	.31	.05
	RG	5.5	19.5	25.0	.45	.07
	NB	1.0	1.0	2.0	.03	.02
	NG	0.8	5.5	6.3	.12	.04

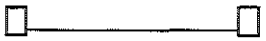
Note: [a] RB = Referred Boys; RG = Referred Girls; NB = Nonreferred Boys; NG = Nonreferred Girls; [b] SE = Standard error of the mean

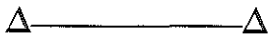
Appendix E.4
Percentage of CBCL/2-3 items (scored 1 or 2) for matched referred and nonreferred children grouped according to referral status, sex, and age

Legenda of the Figures:

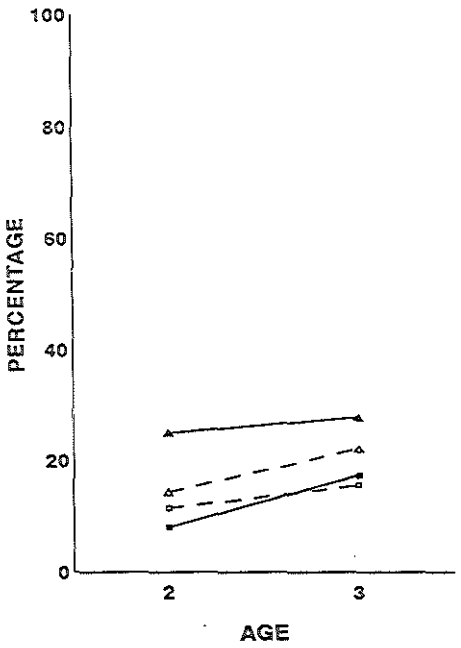

Referred Boys


Referred Girls

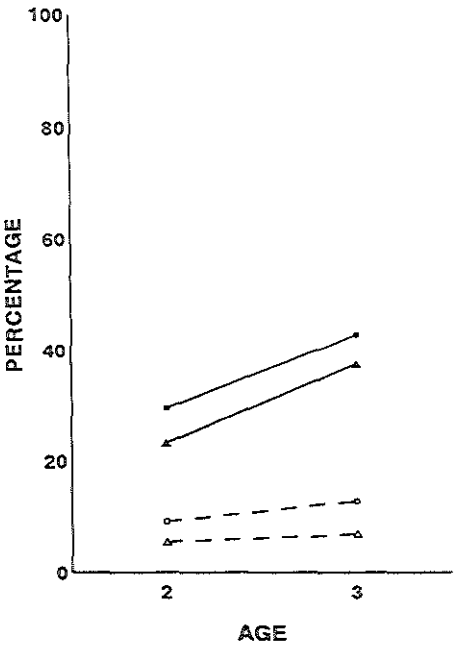

Nonreferred Boys


Nonreferred Girls

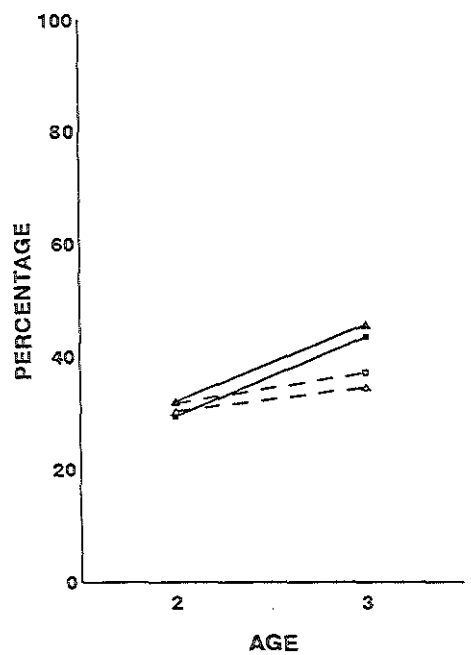
1. ACHES



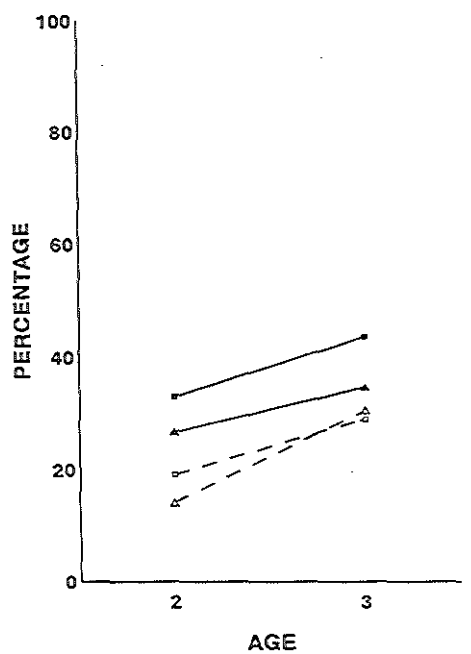
2. ACTS TOO YOUNG



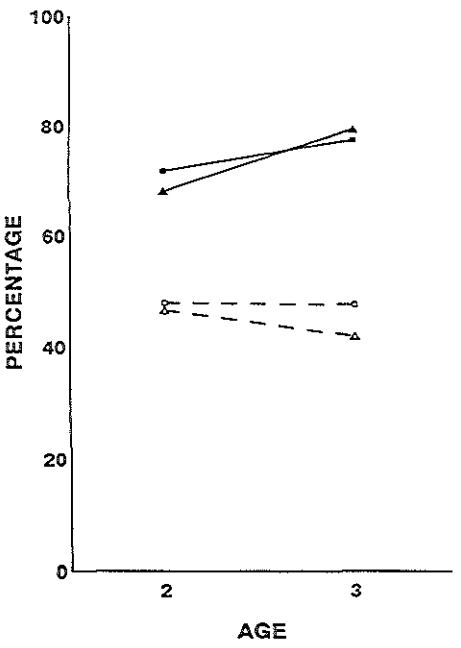
3. AFRAID TOO TRY NEW THINGS



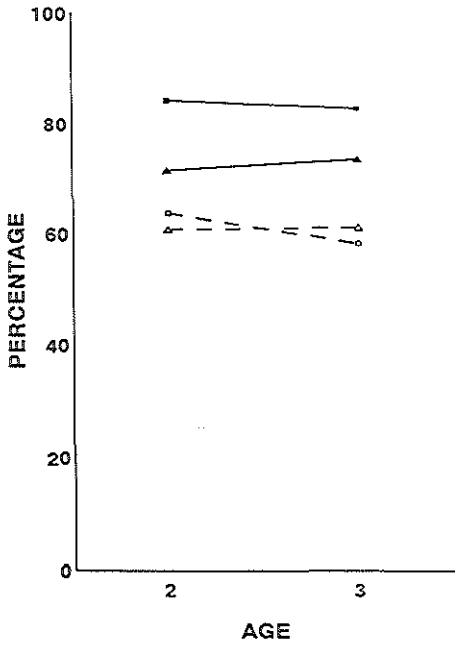
4. AVOIDS EYE CONTACT



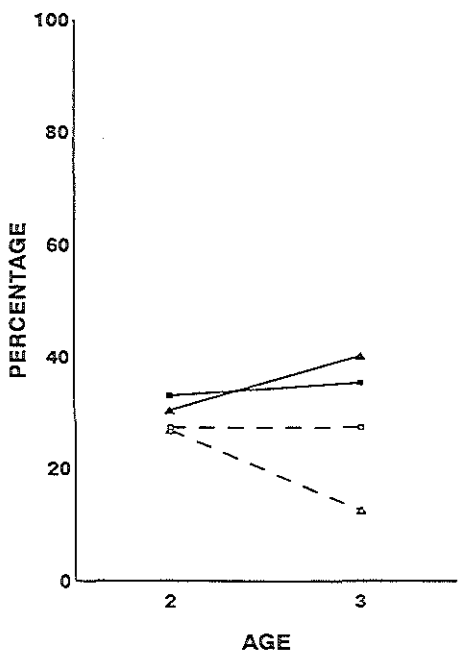
5. CAN'T CONCENTRATE



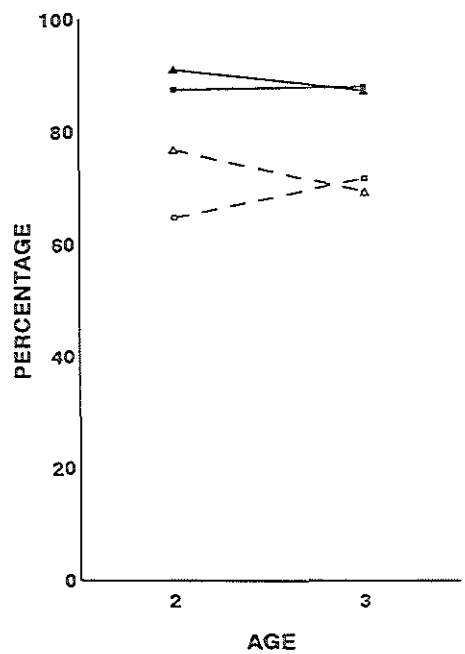
6. CAN'T SIT STILL



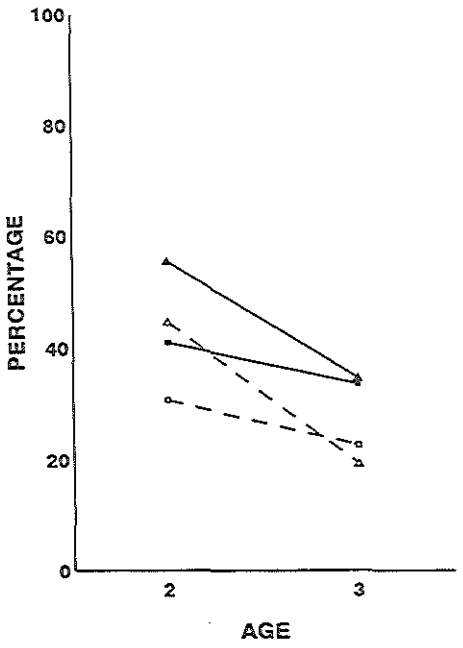
7. CAN'T STAND THINGS MOVED



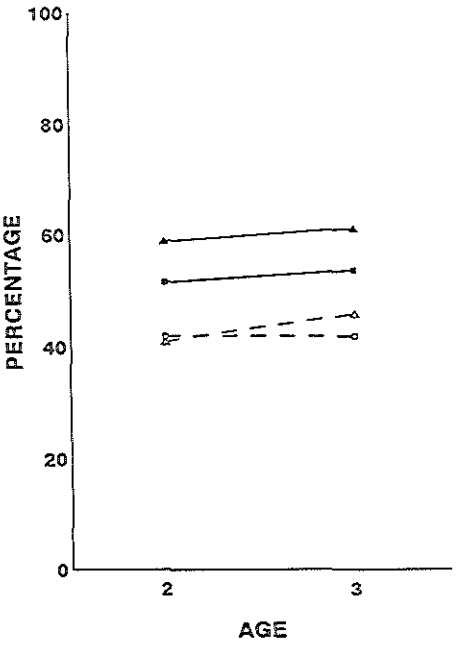
8. CAN'T WAIT



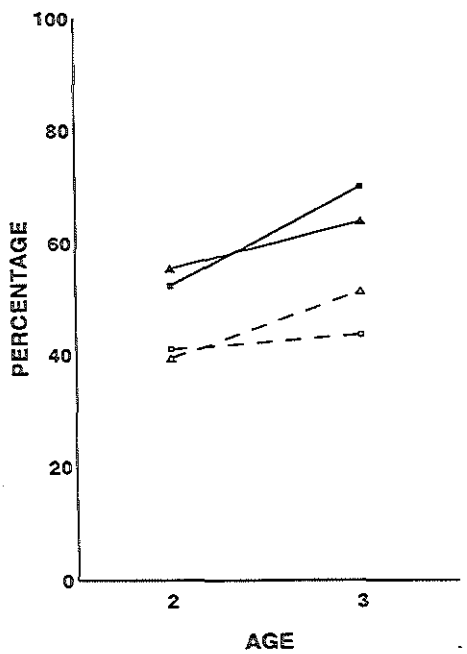
9. CHEWS NONFOOD



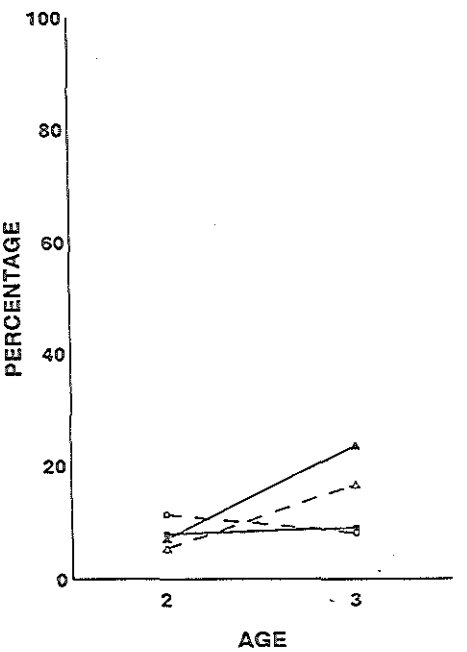
10. CLINGS TO ADULTS



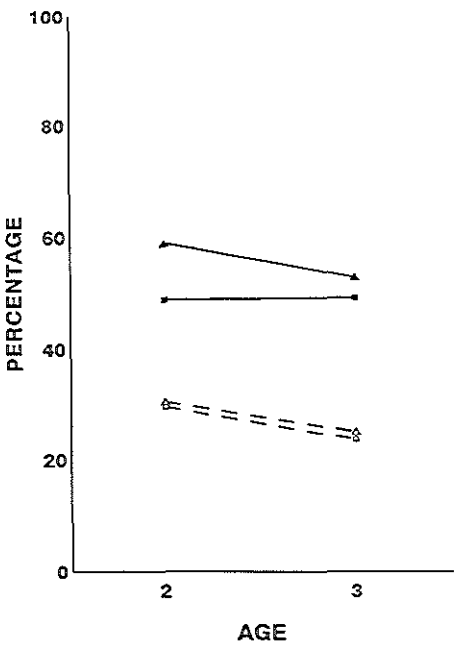
11. CONSTANTLY SEEKS HELP



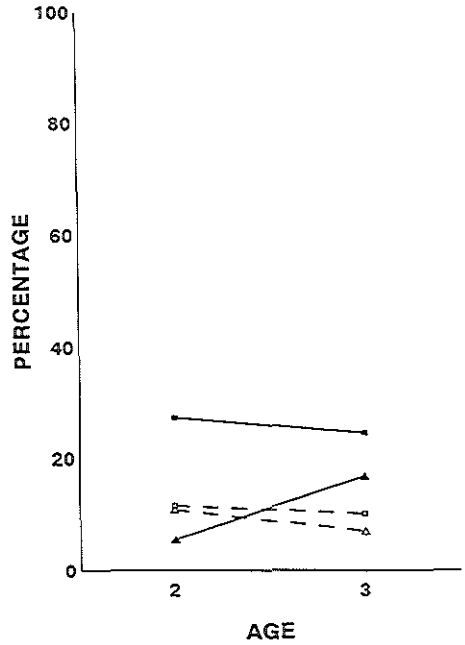
12. CONSTIPATED



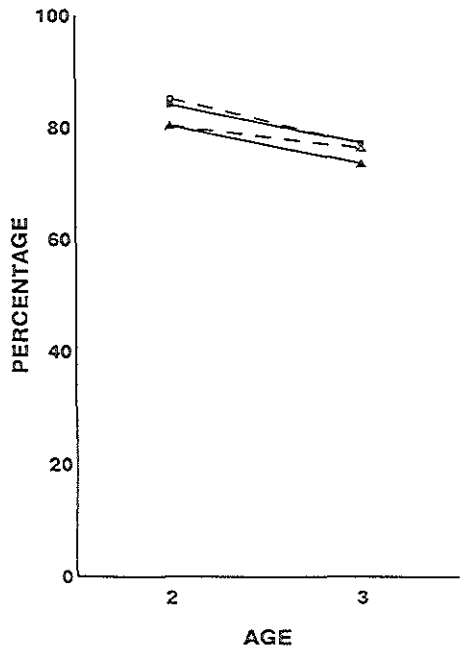
13. CRIES MUCH



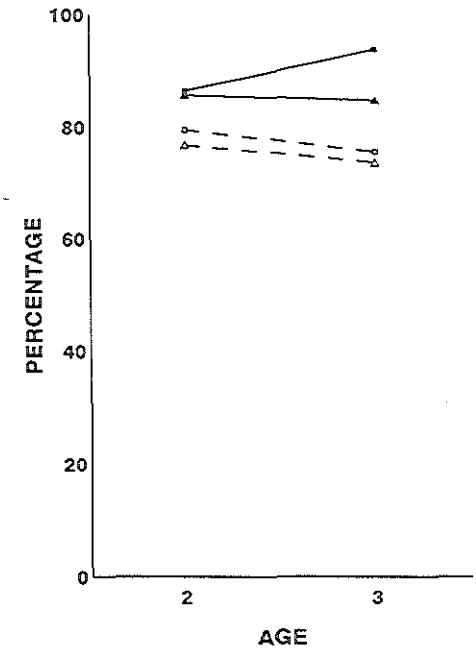
14. CRUEL TO ANIMALS



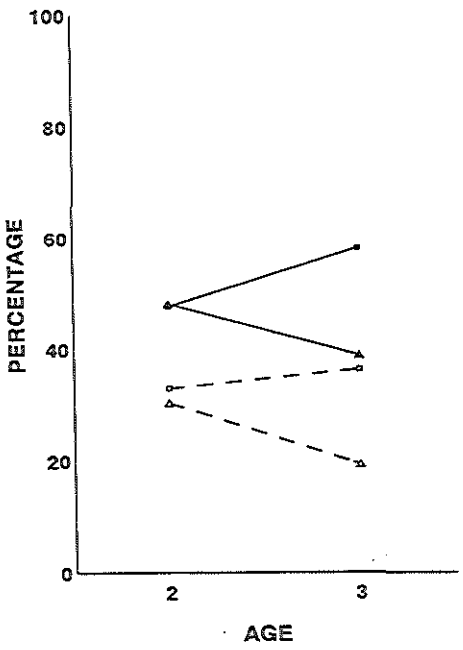
15. DEFIANT



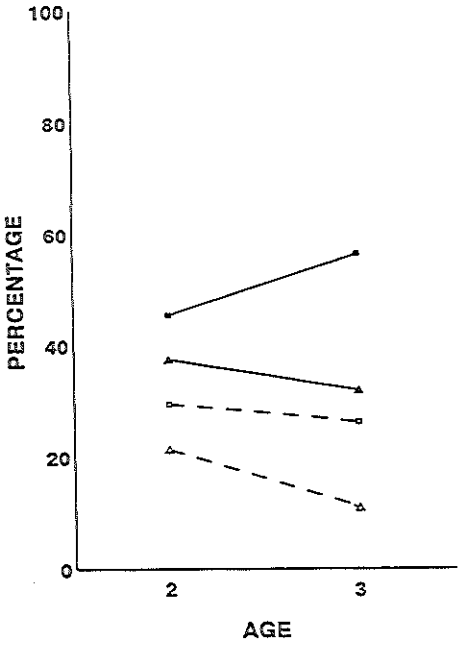
16. DEMANDS MUST BE MET



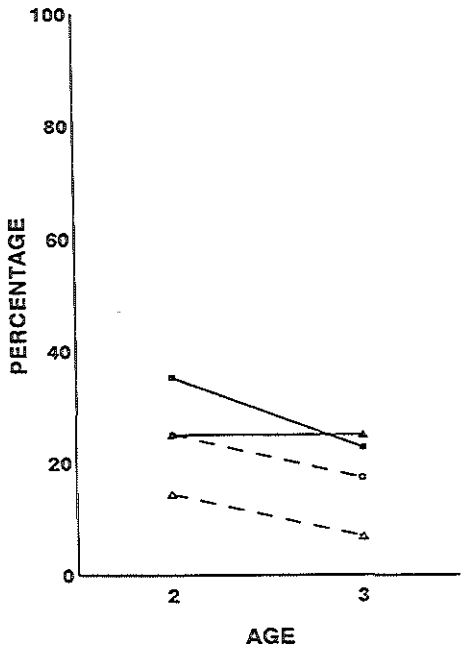
17. DESTROYS OWN THINGS



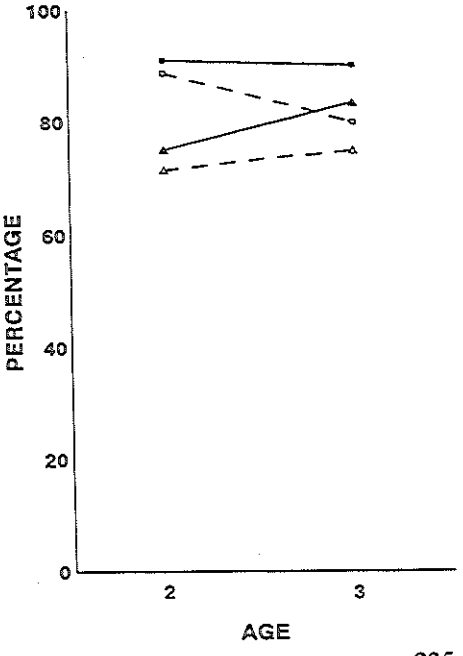
18. DESTROYS OTHER'S THINGS



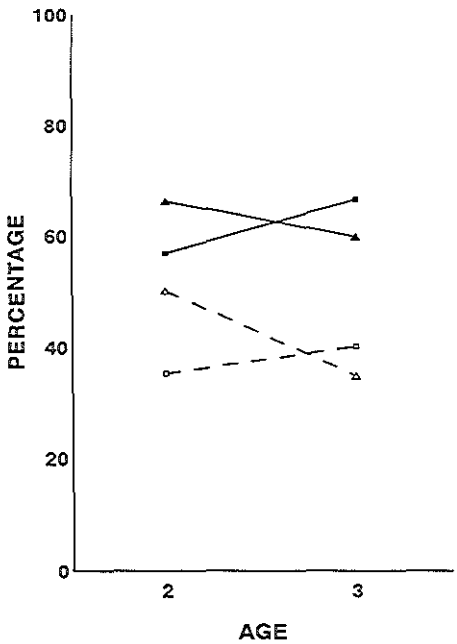
19. DIARRHEA



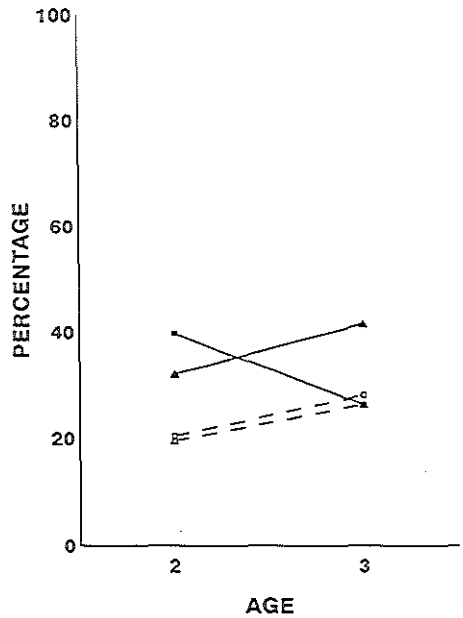
20. DISOBEDIENT



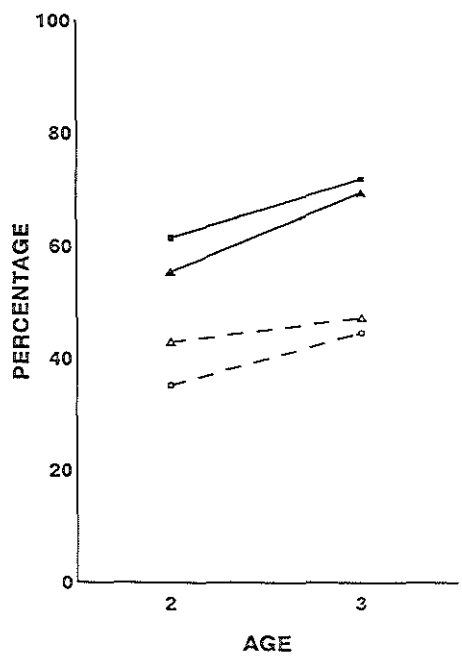
21. DISTURBED BY CHANGE



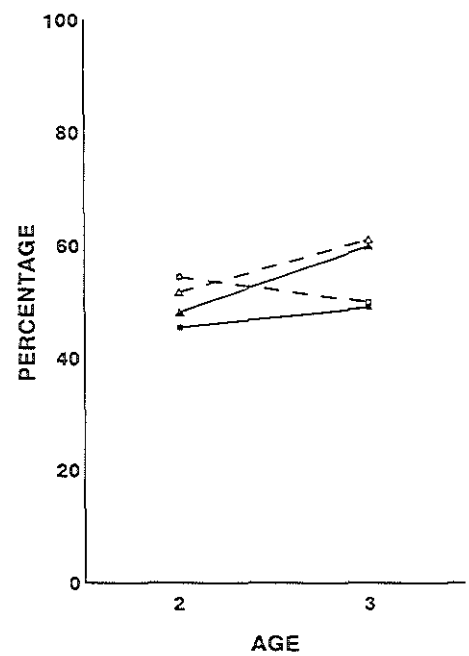
22. DOESN'T WANT TO SLEEP ALONE



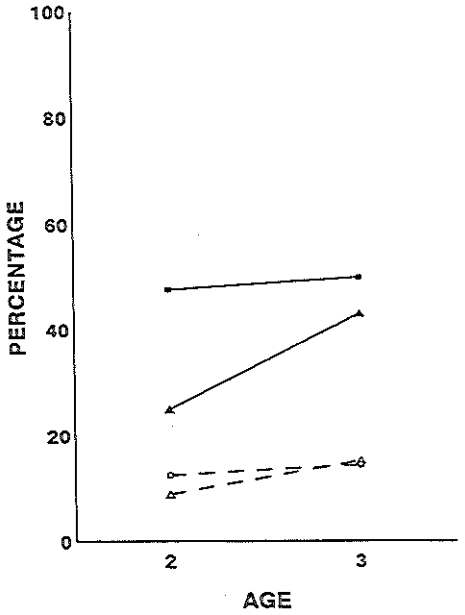
23. DOESN'T ANSWER



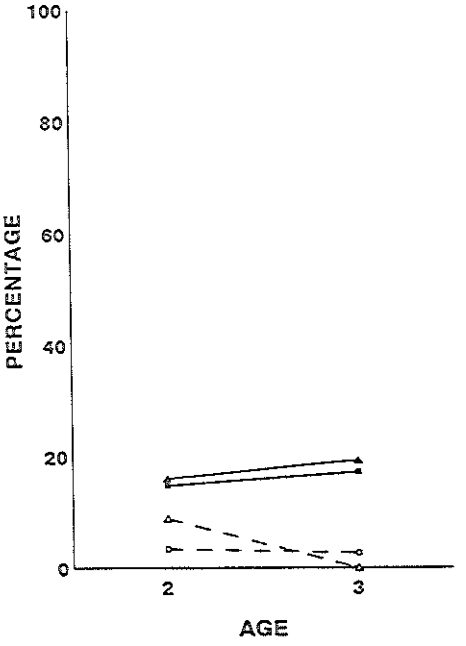
24. DOESN'T EAT WELL



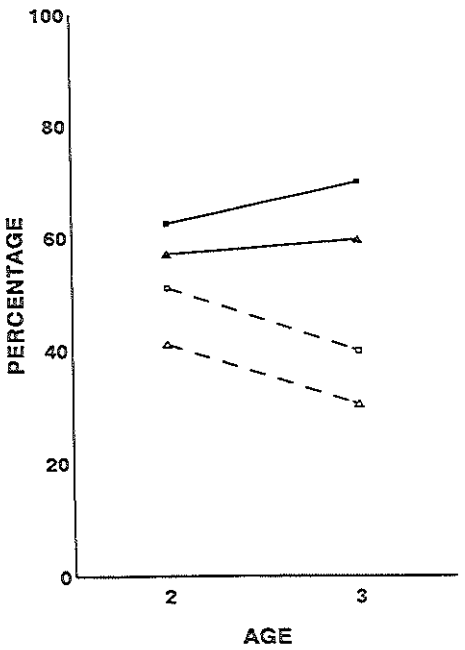
25. DOESN'T GET ALONG WITH OTHER KIDS



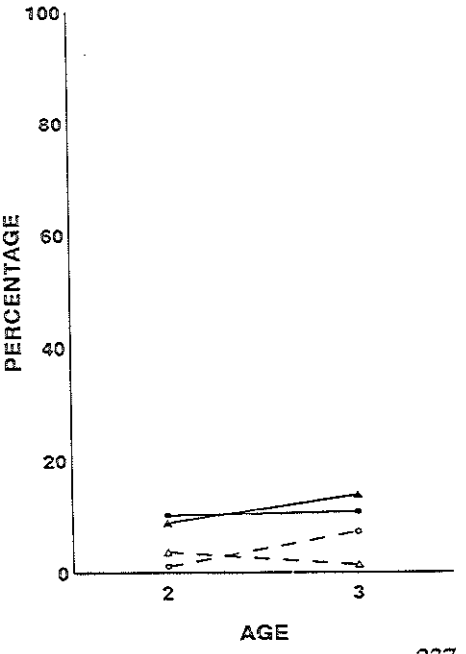
26. NO FUN



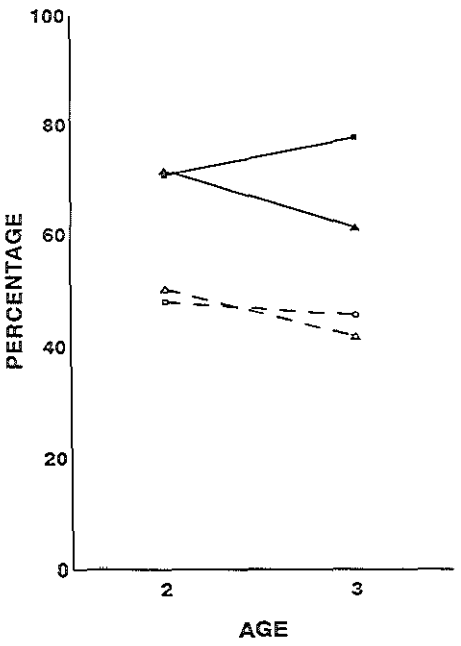
27. LACKS GUILT



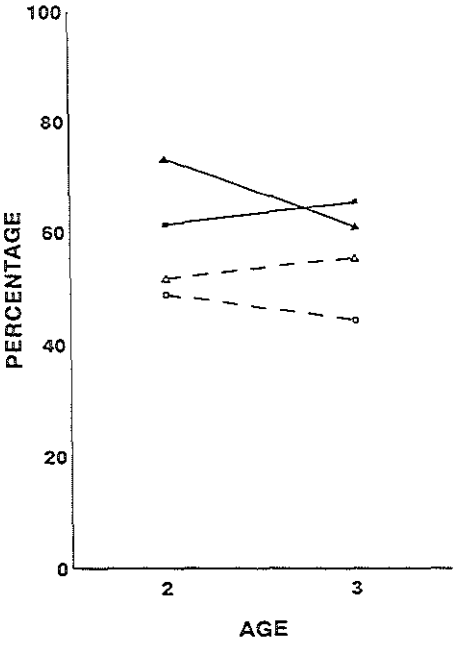
28. DOESN'T WANT TO GO OUT



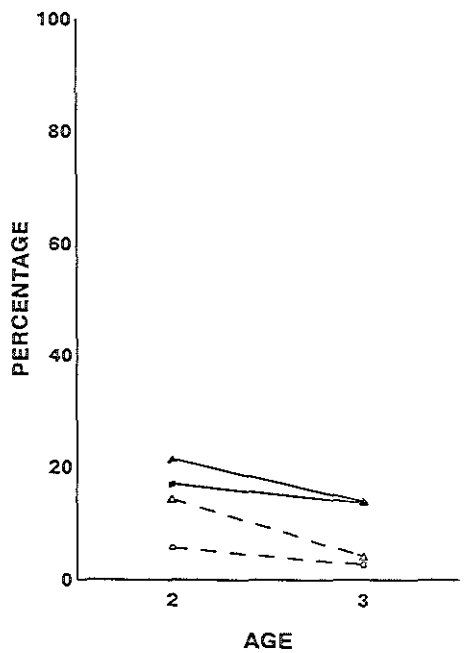
29. EASILY FRUSTRATED



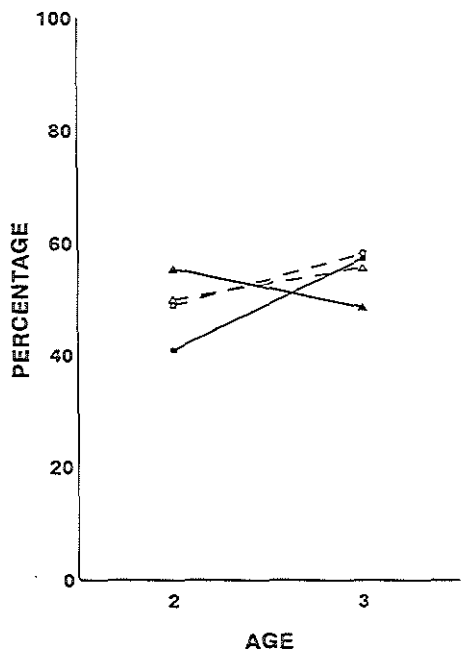
30. EASILY JEALOUS



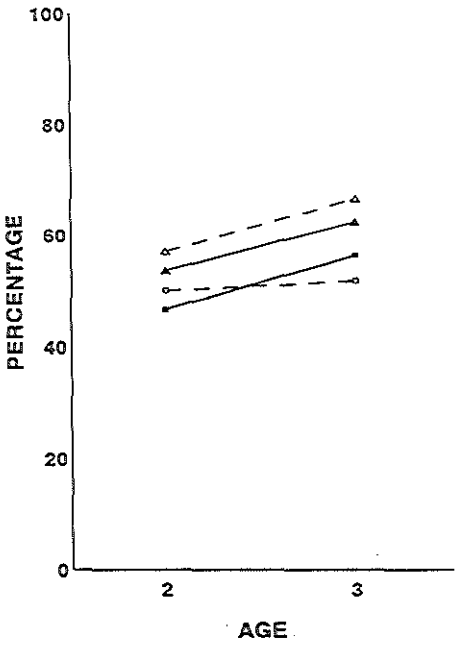
31. EATS OR DRINKS NONFOOD



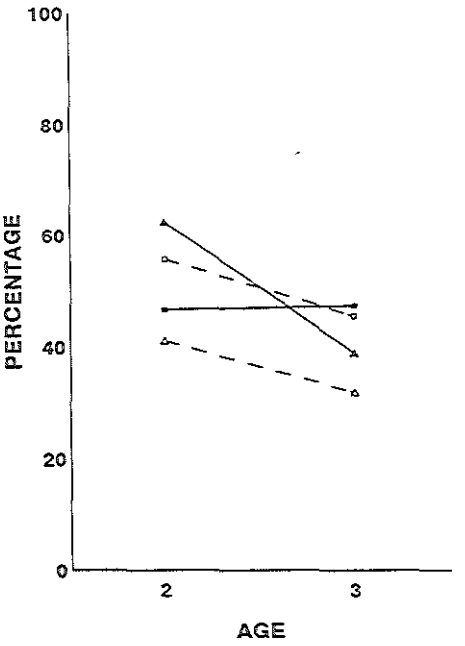
32. FEARS



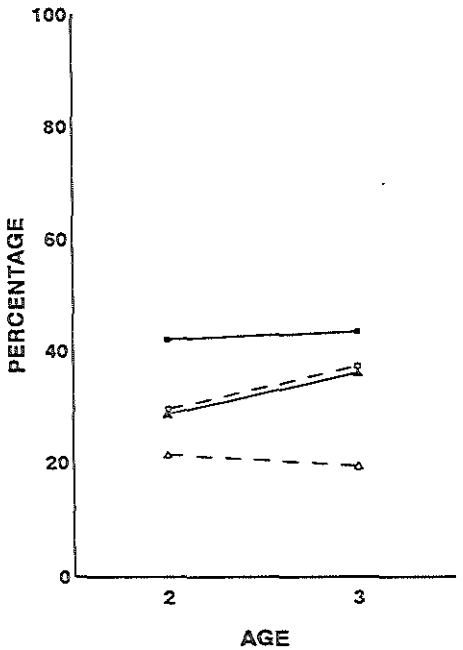
33. FEELINGS ARE EASILY HURT



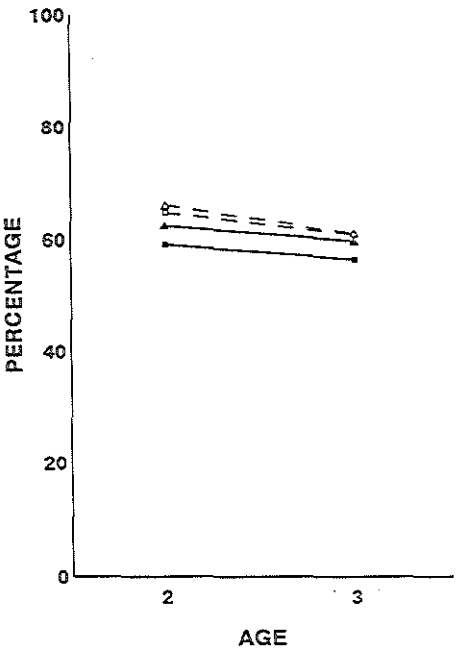
34. ACCIDENT-PRONE



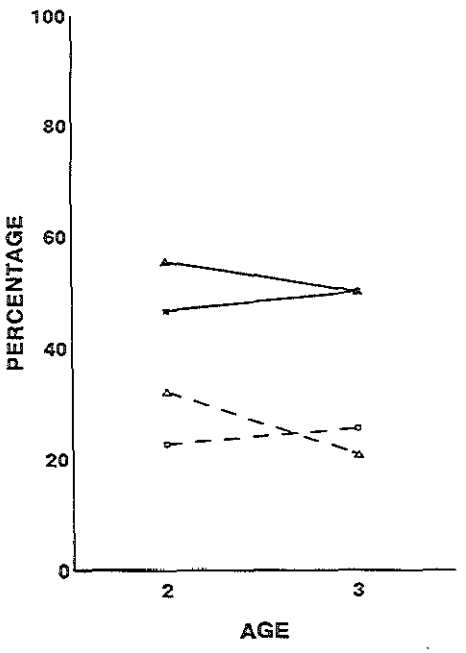
35. FIGHTS



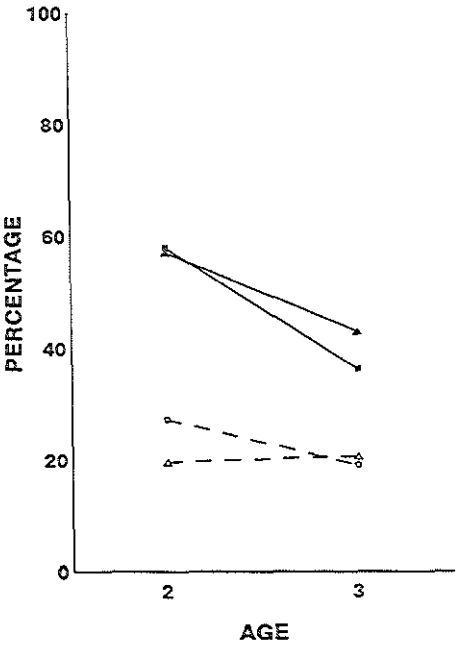
36. GETS INTO EVERYTHING



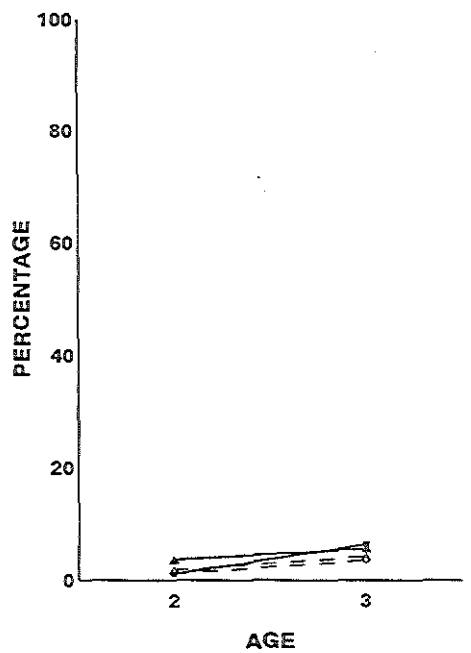
37. UPSET BY SEPARATION



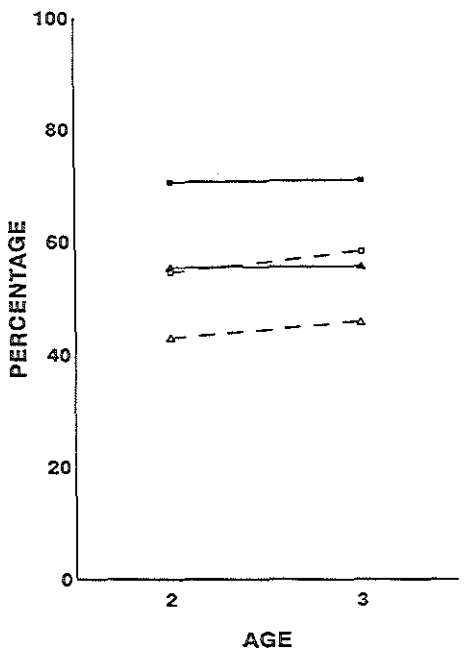
38. CAN'T SLEEP



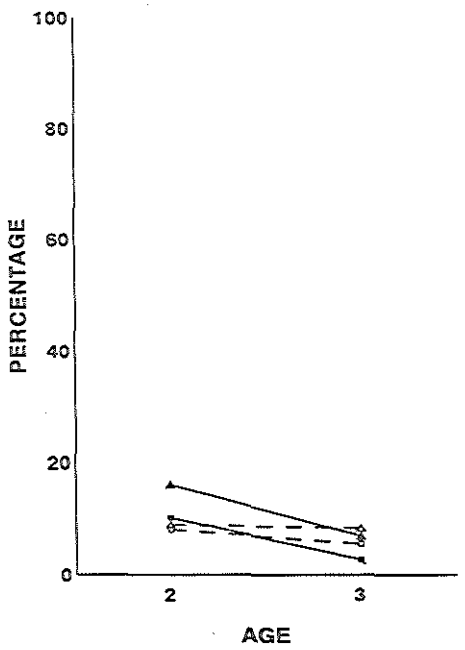
39. HEADACHES



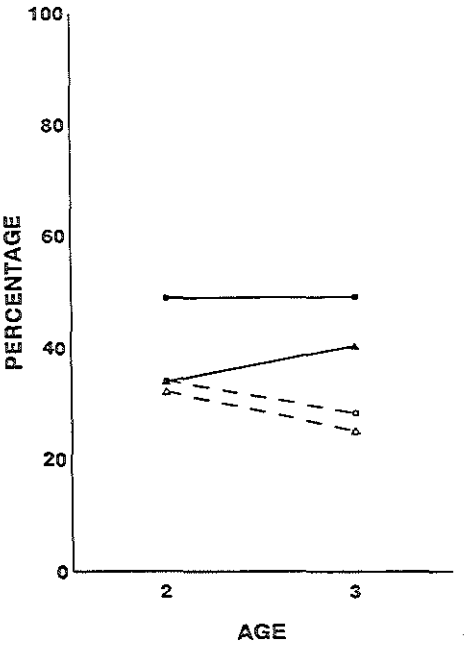
40. HITS



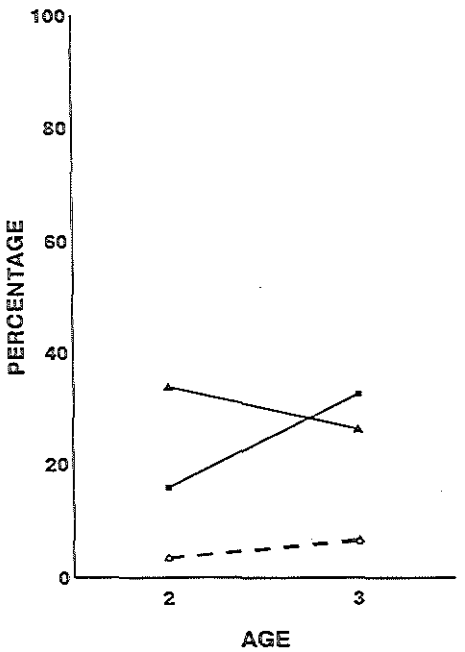
41. HOLDS BREATH



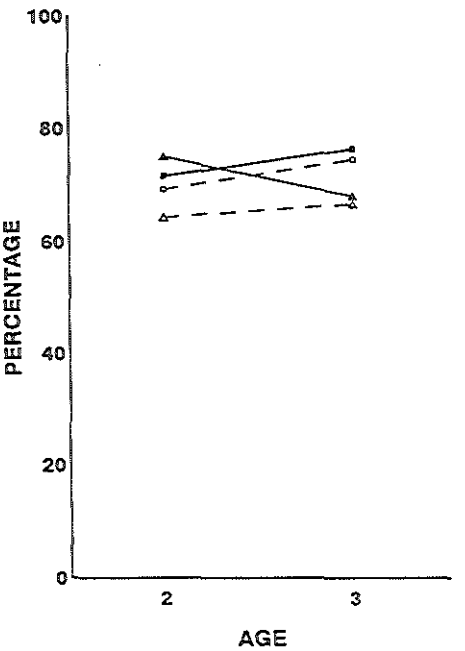
42. HURTS ACCIDENTALLY



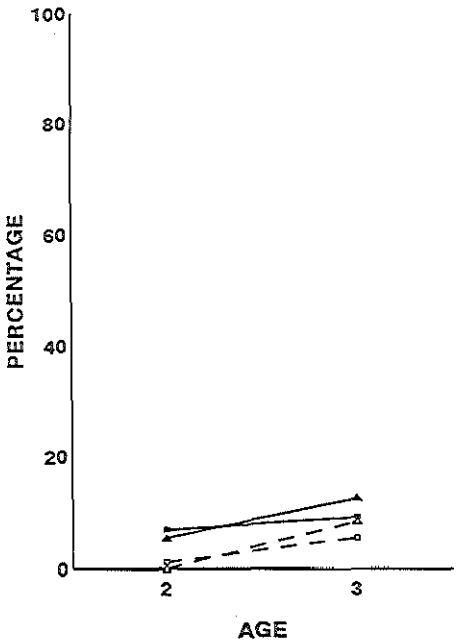
43. LOOKS UNHAPPY



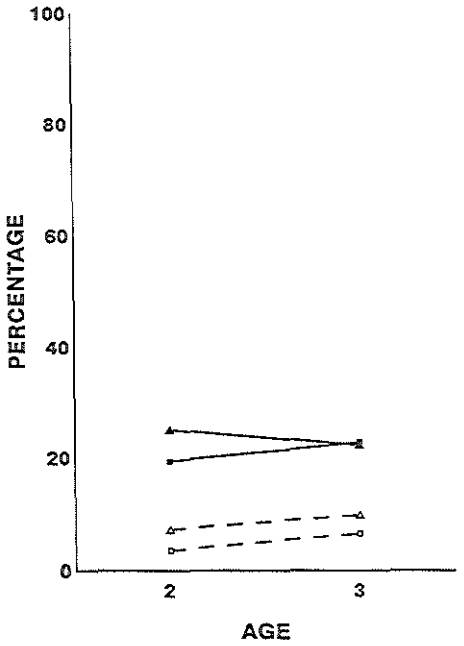
44. ANGRY MOODS



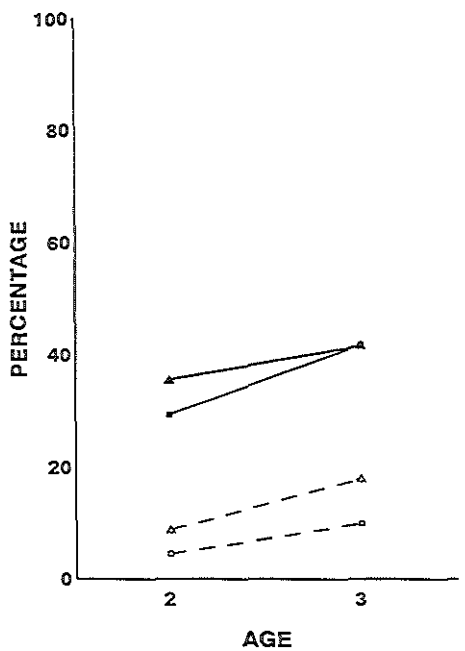
45. NAUSEA



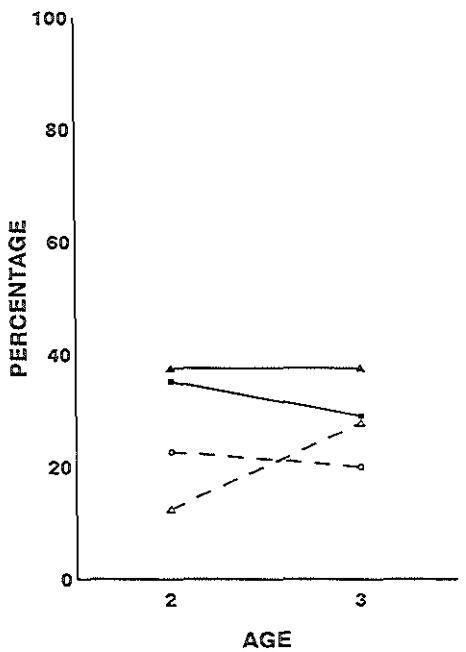
46. NERVOUS MOVEMENTS



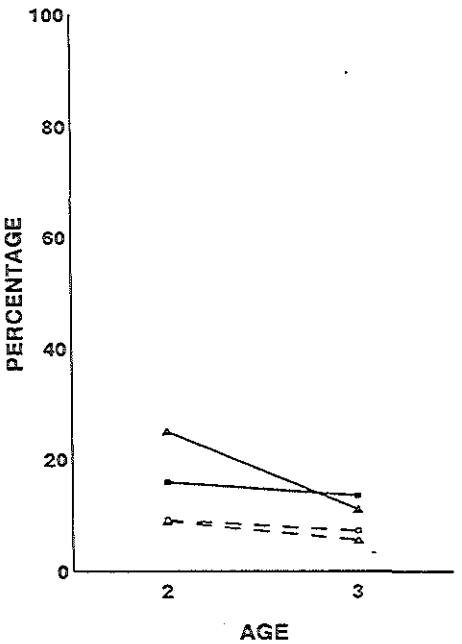
47. NERVOUS



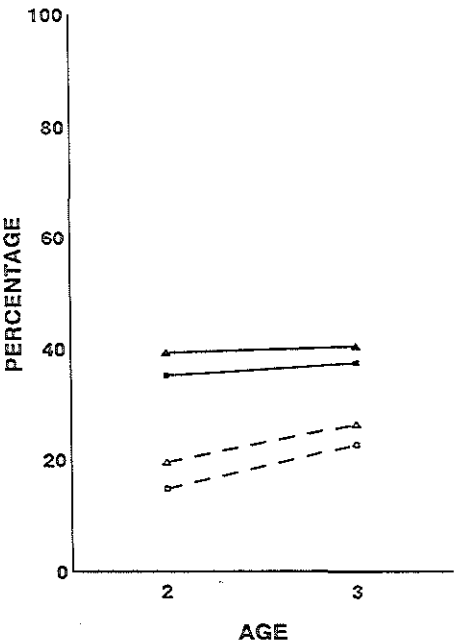
48. NIGHTMARES



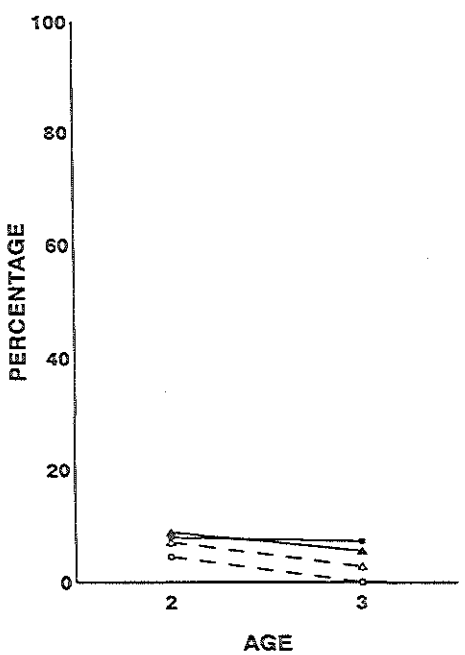
49. OVEREATING



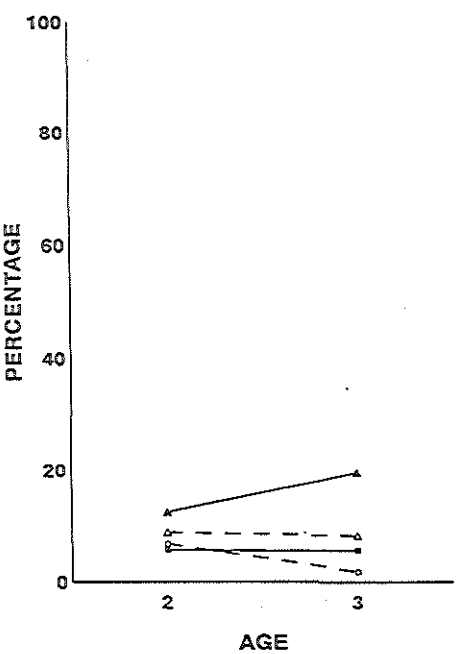
50. OVERTIRED



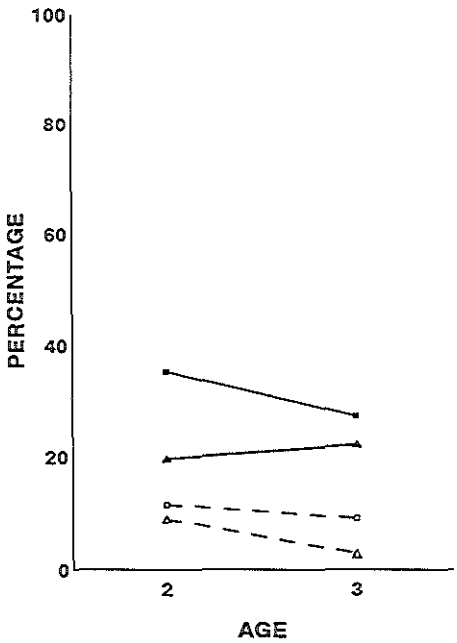
51. OVERWEIGHT



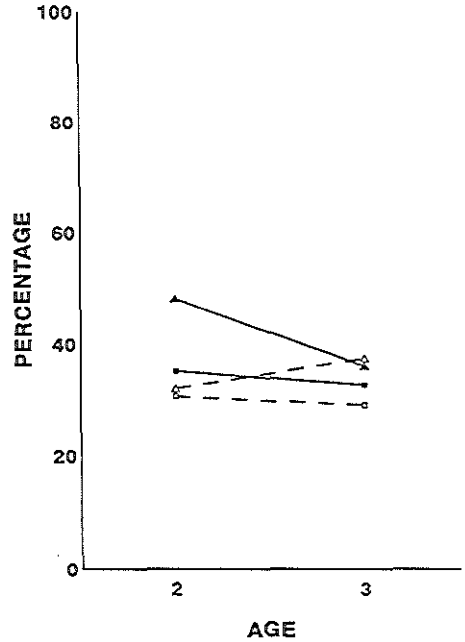
52. PAINFUL BOWEL MOVEMENTS



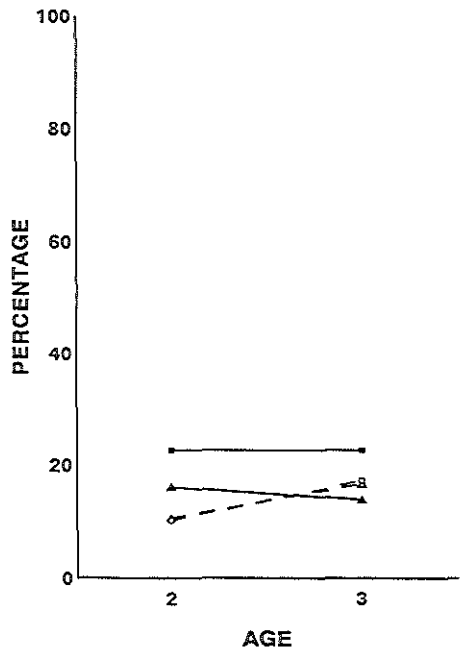
53. ATTACKS PEOPLE



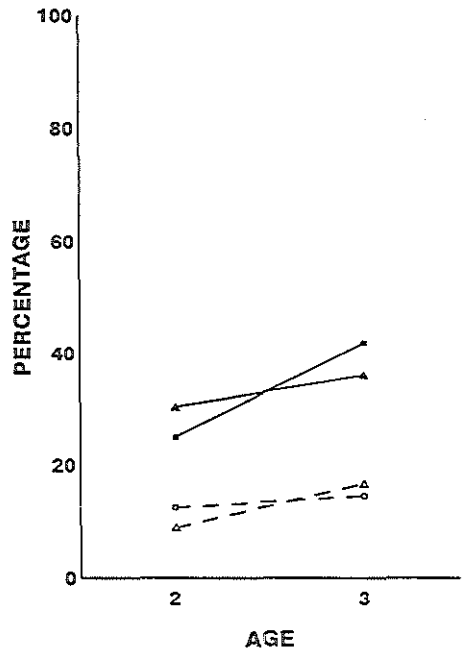
54. PICKING



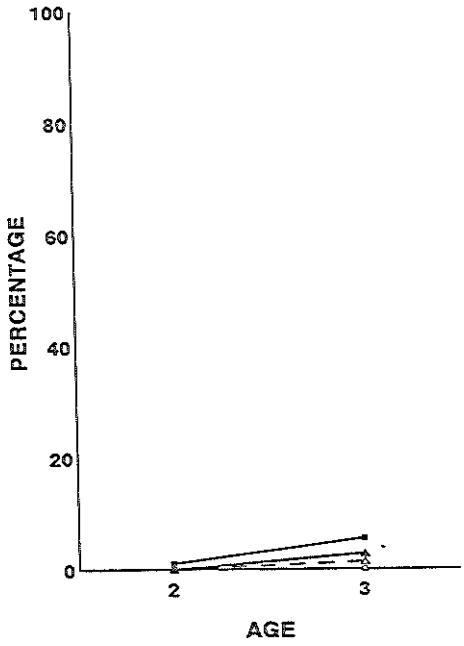
55. PLAYS WITH SEX PARTS



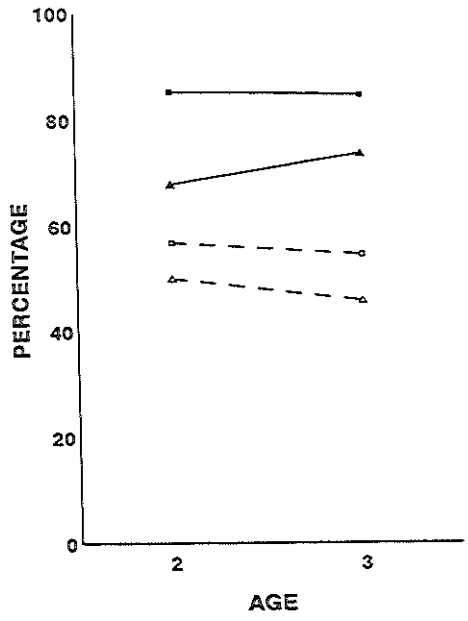
56. CLUMSY



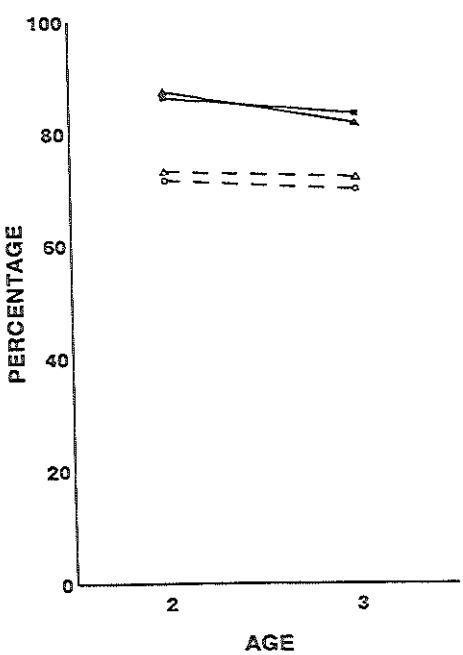
57. EYE PROBLEMS



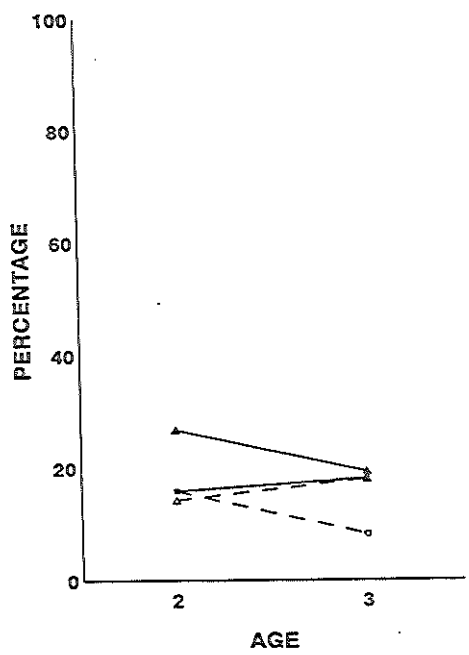
58. PUNISHMENT DOESN'T CHANGE BEHAVIOR



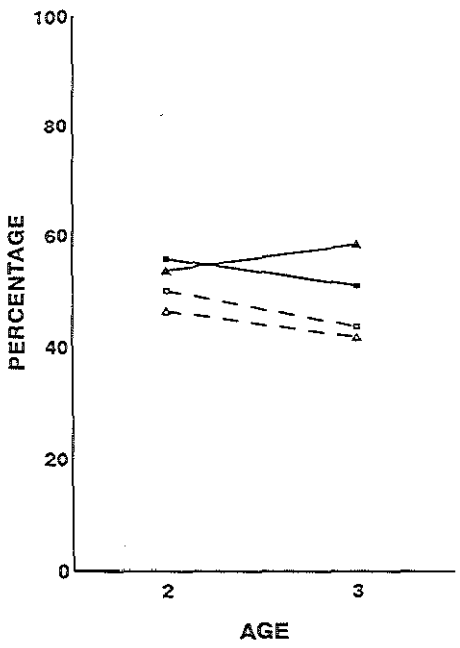
59. QUICKLY SHIFTS ACTIVITY



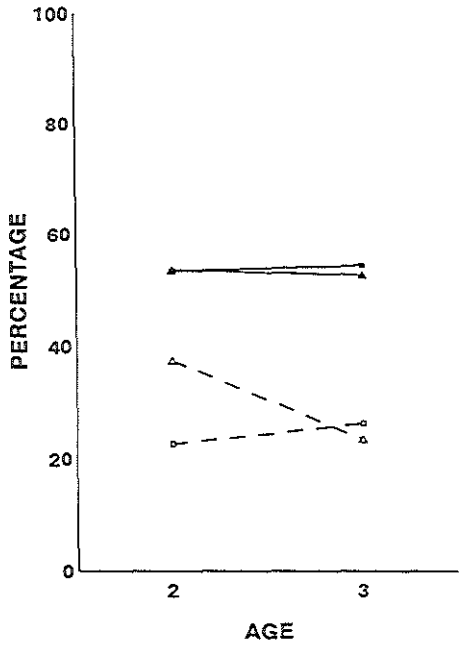
60. RASHES



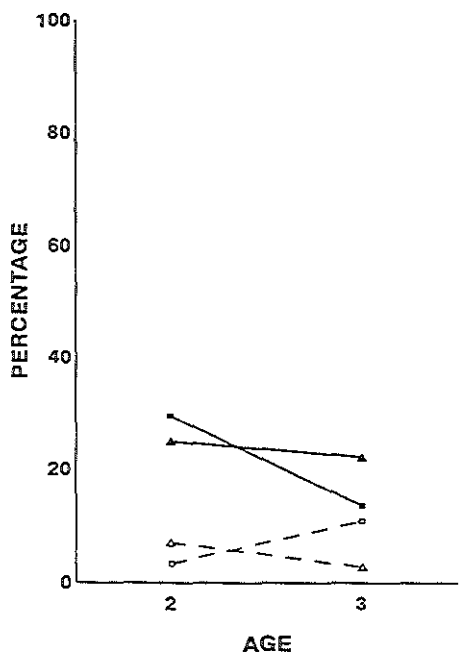
61. REFUSES TO EAT



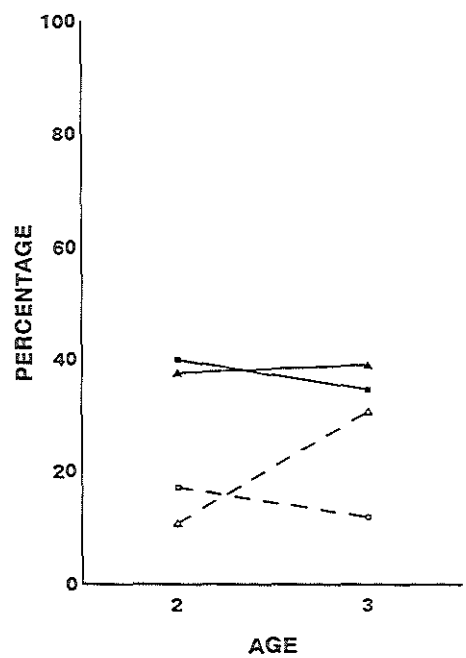
62. REFUSES ACTIVE GAMES



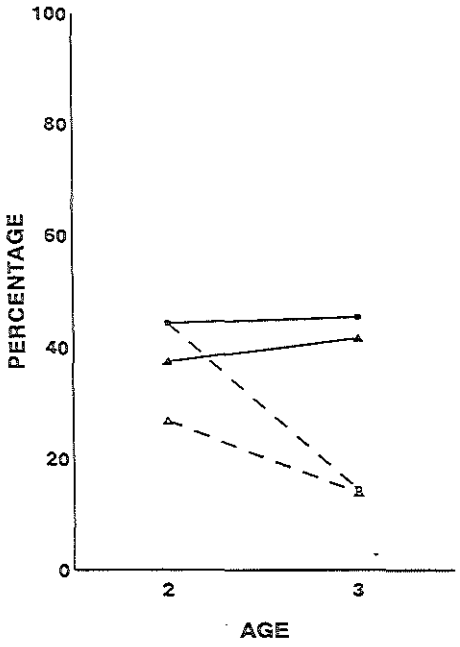
63. ROCKS HEAD OR BODY



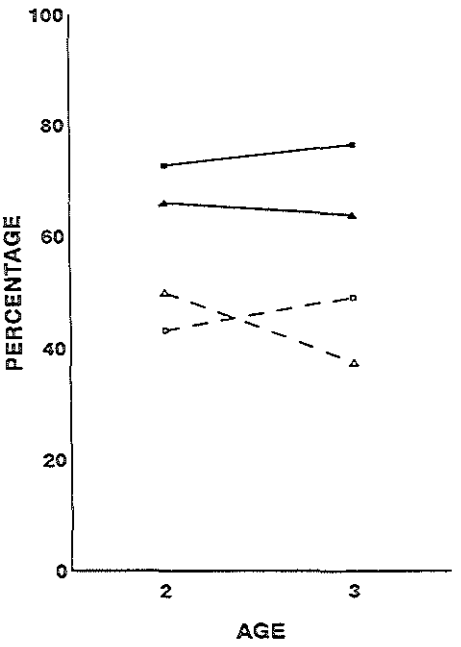
64. RESISTS GOING TO BED



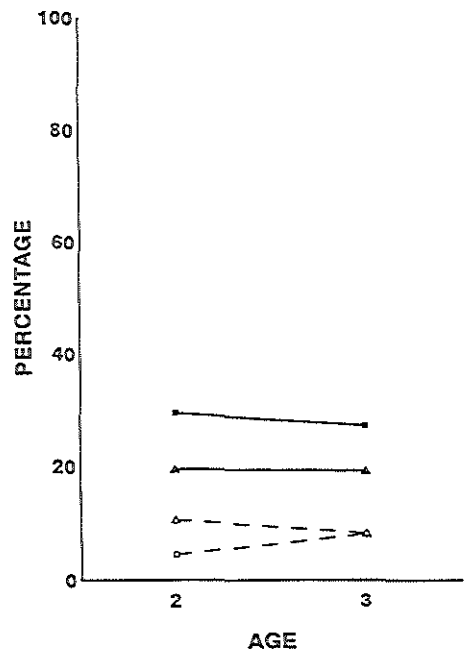
65. RESISTS TOILET TRAINING



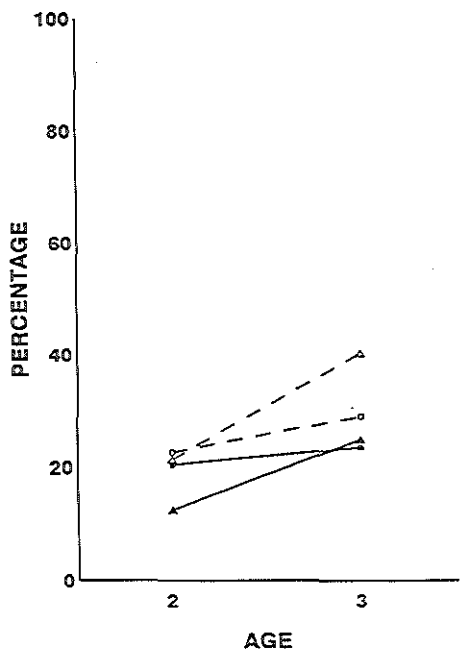
66. SCREAMS



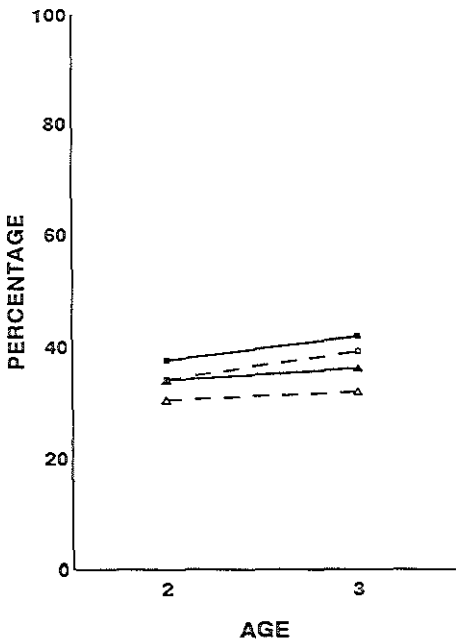
67. UNRESPONSIVE TO AFFECTION



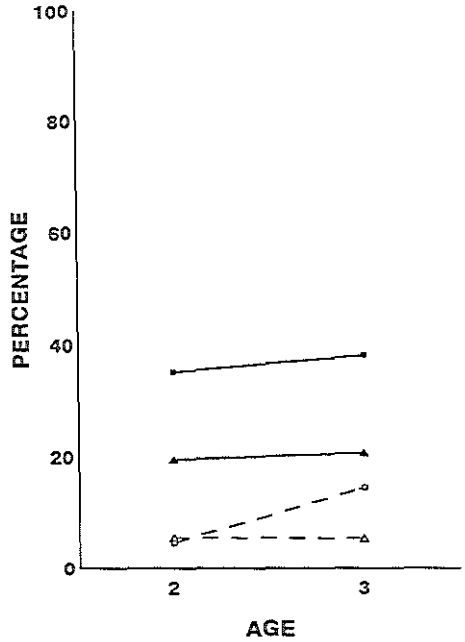
68. SELF-CONSCIOUS



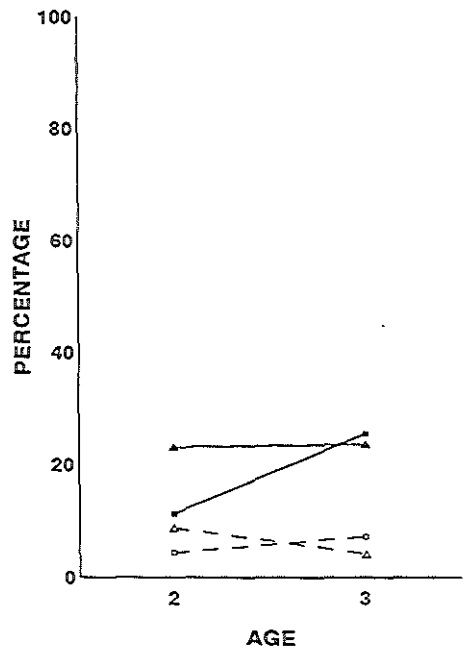
69. SELFISH



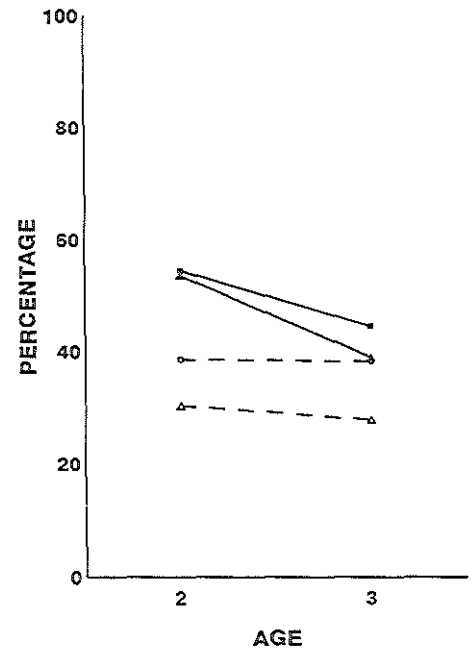
70. LITTLE AFFECTION



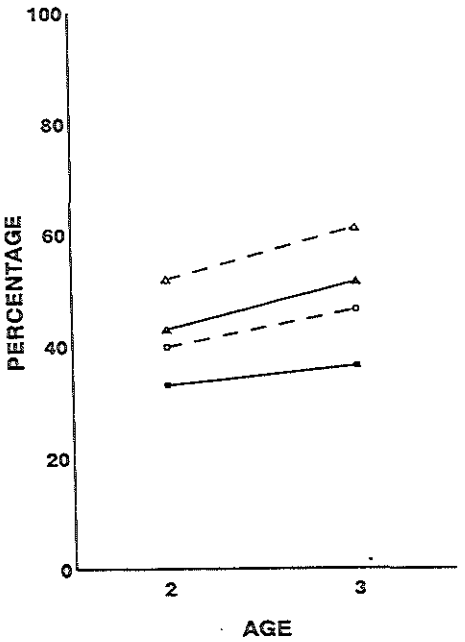
71. LITTLE INTEREST



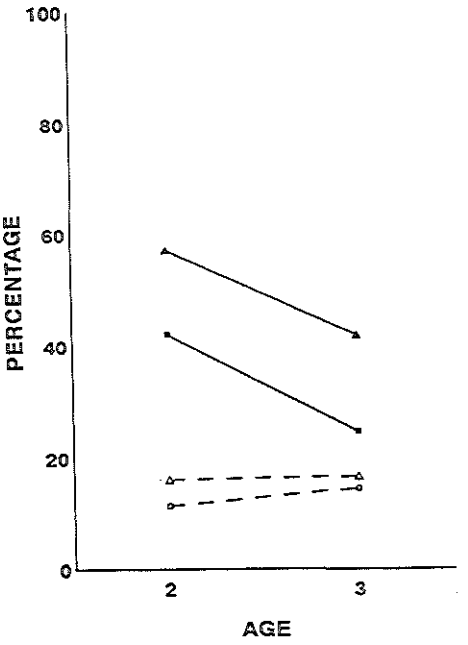
72. LITTLE FEAR



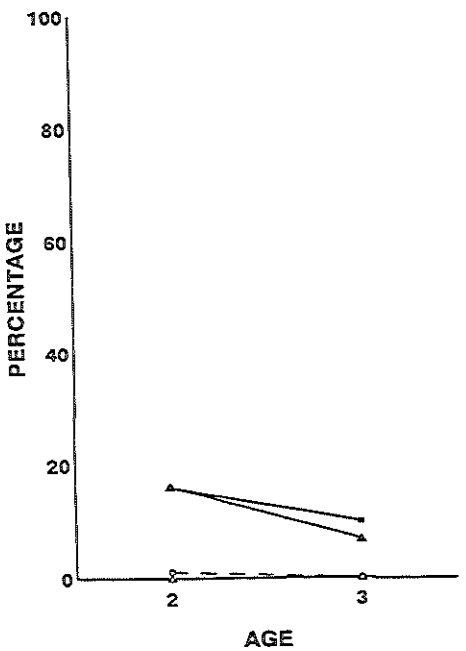
73. SHY



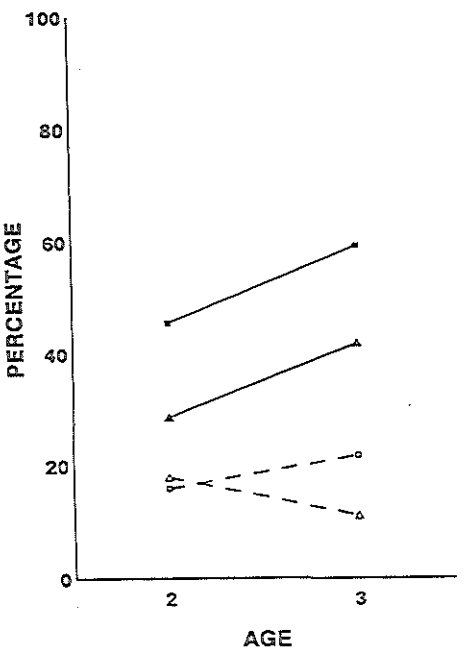
74. SLEEPS LITTLE



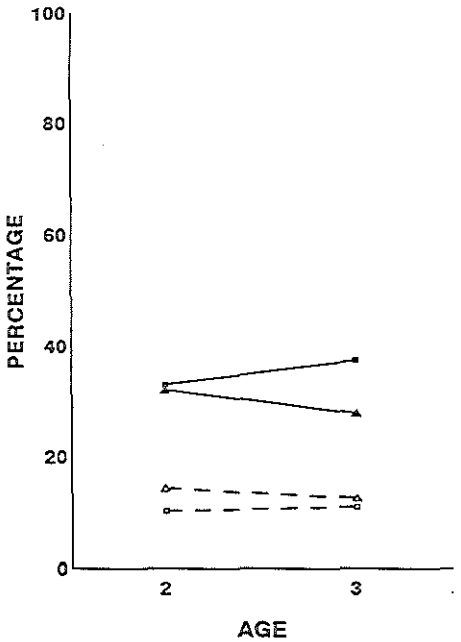
75. SMEARS BOWEL MOVEMENTS



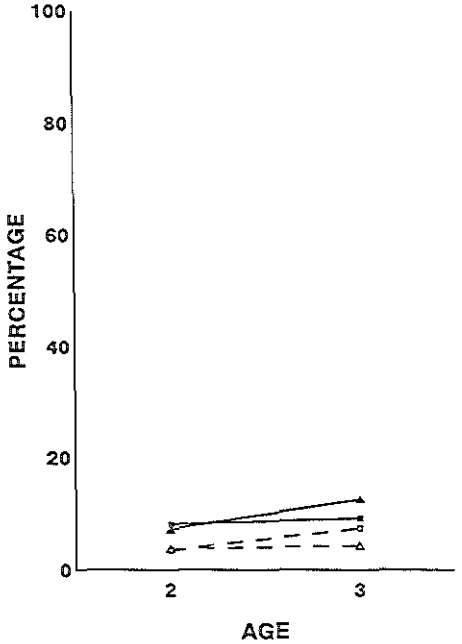
76. SPEECH PROBLEM



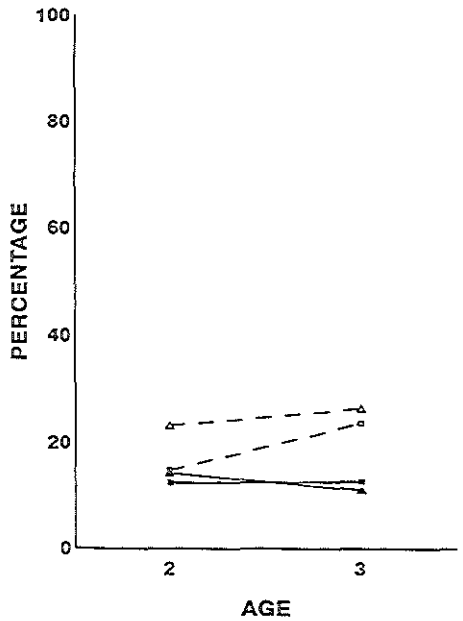
77. STARES BLANKLY



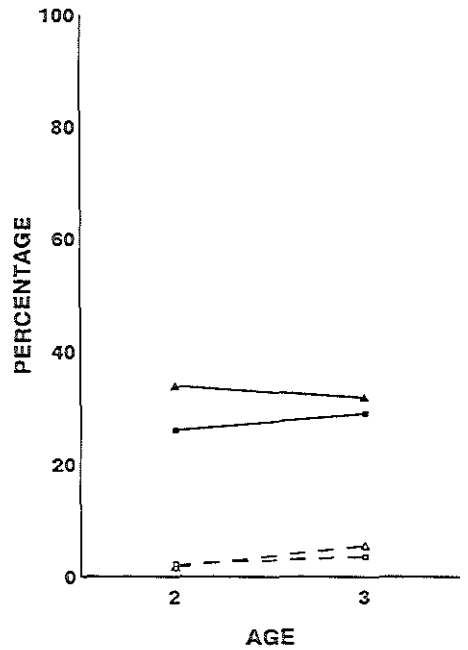
78. STOMACHACHES



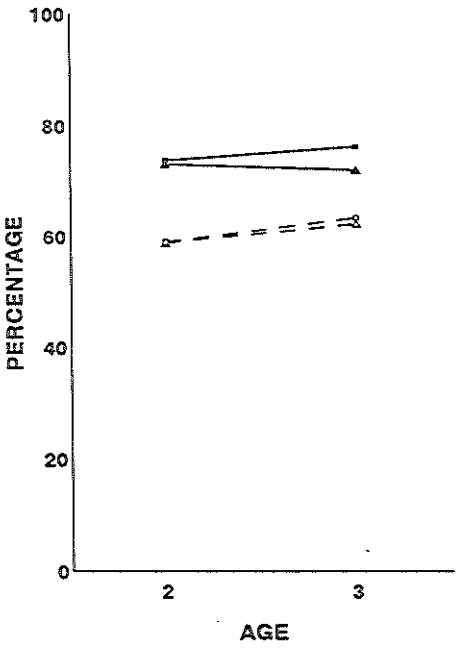
79. STORES UP UNNEEDED THINGS



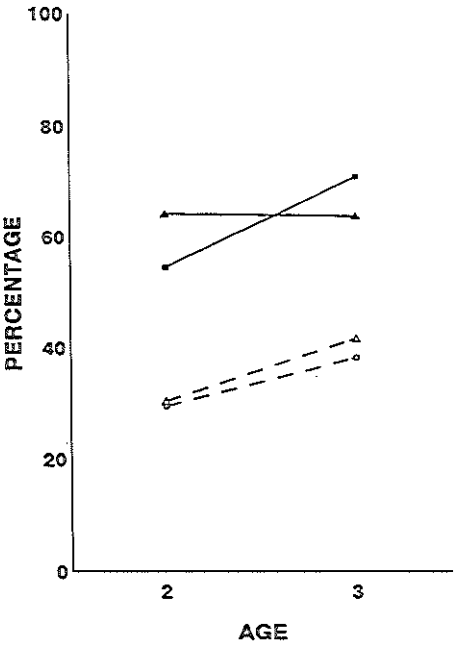
80. STRANGE BEHAVIOR



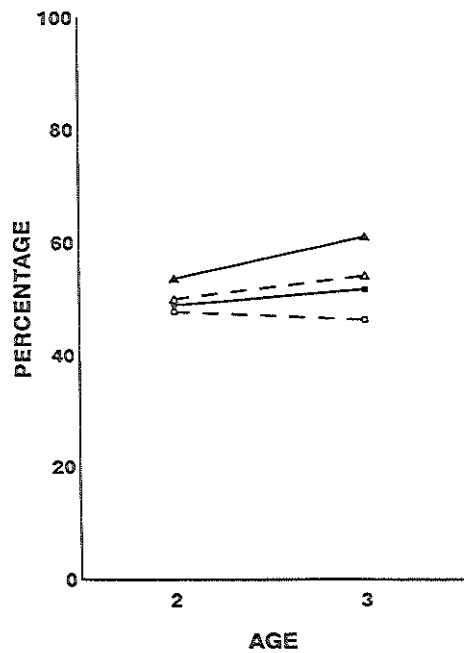
81. STUBBORN



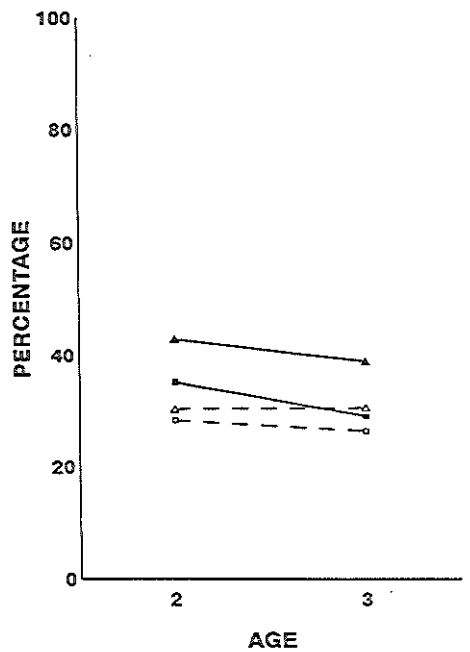
82. MOODY



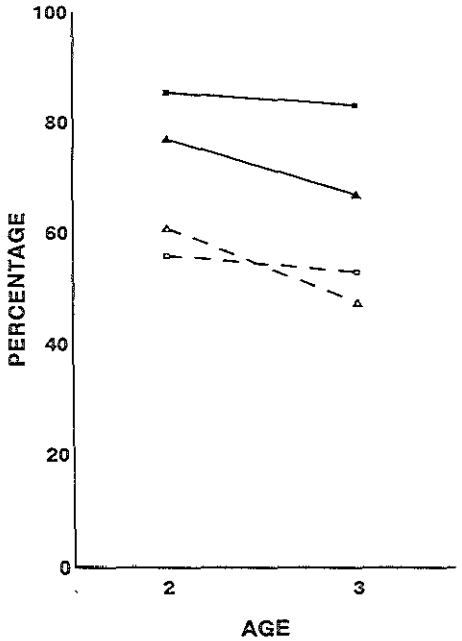
83. SULKS



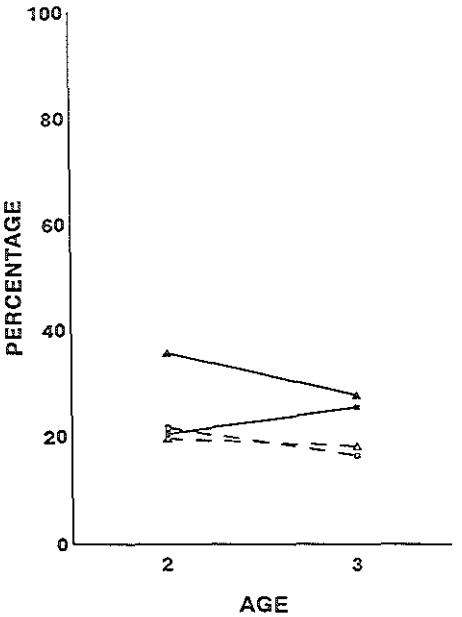
84. TALKS OR CRIES IN SLEEP



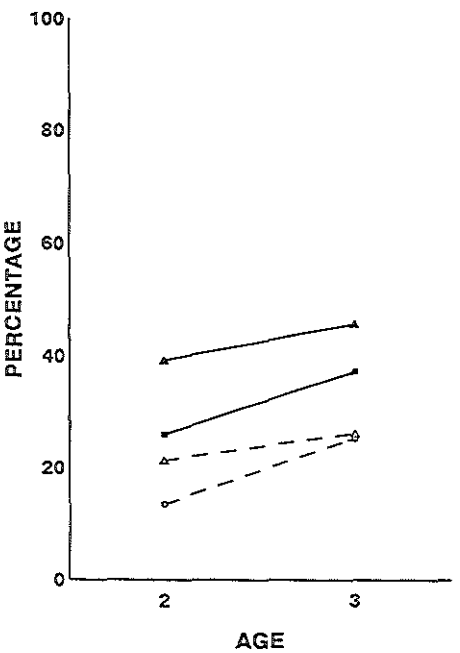
85. TEMPER TANTRUMS



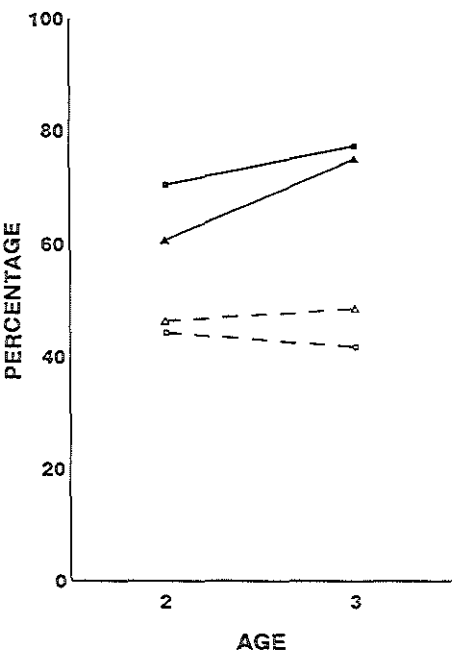
86. TOO CONCERNED WITH NEAT OR CLEAN



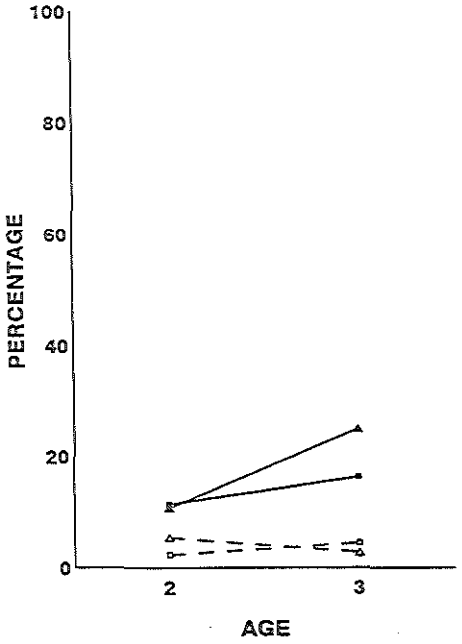
87. TOO FEARFUL OR ANXIOUS



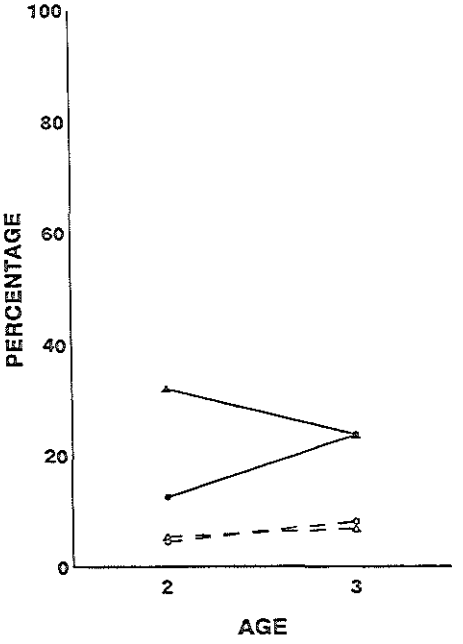
88. UNCOOPERATIVE



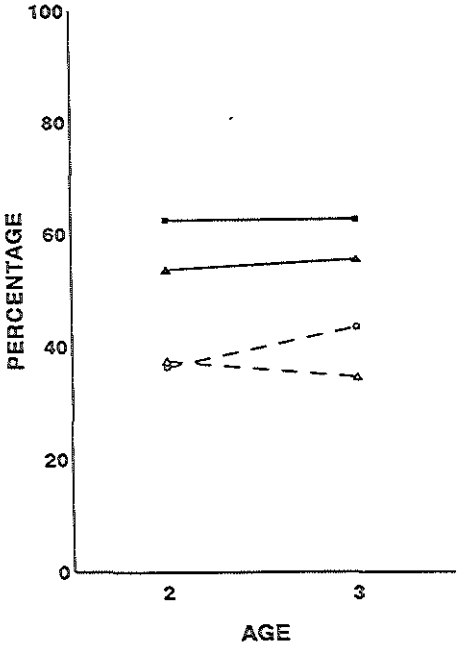
89. UNDERACTIVE



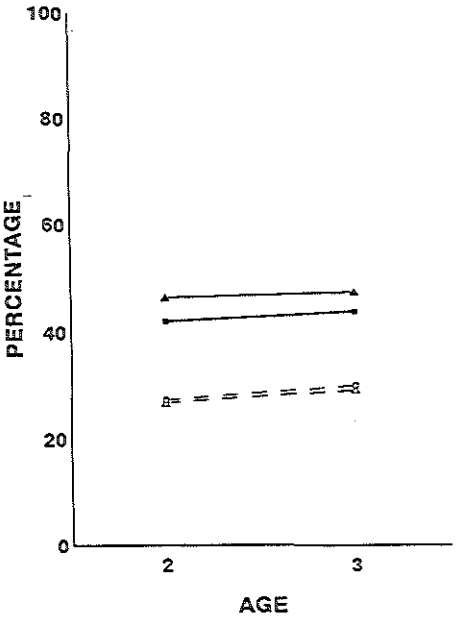
90. SAD



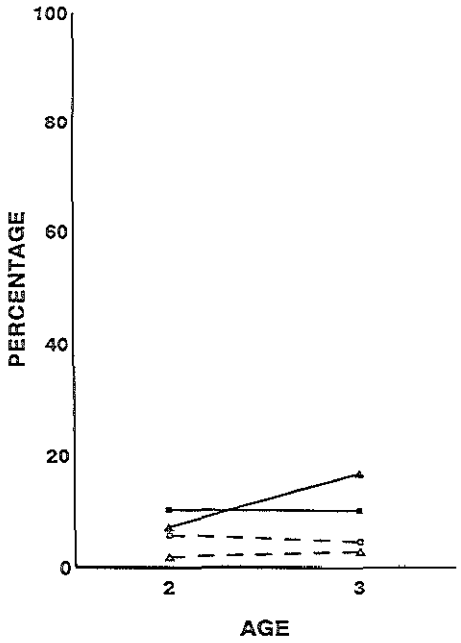
91. TOO LOUD



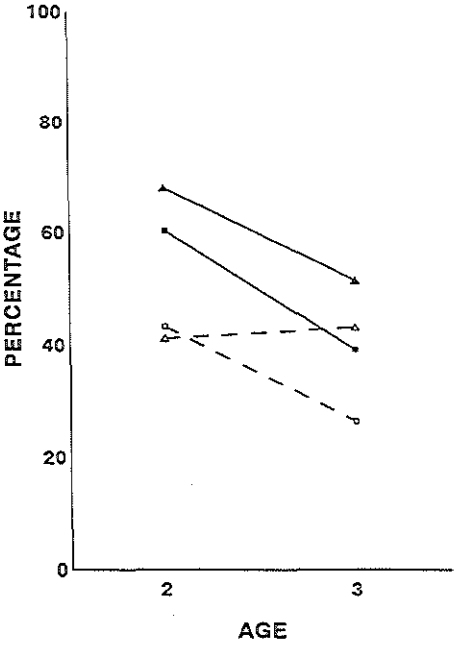
92. UPSET BY NEW PEOPLE OR SITUATIONS



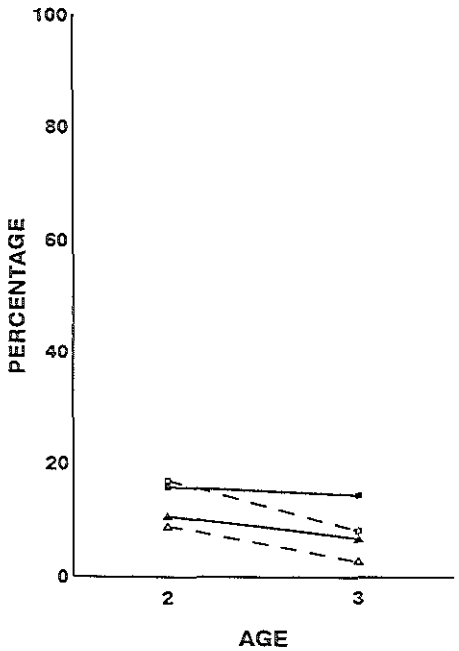
93. VOMITING



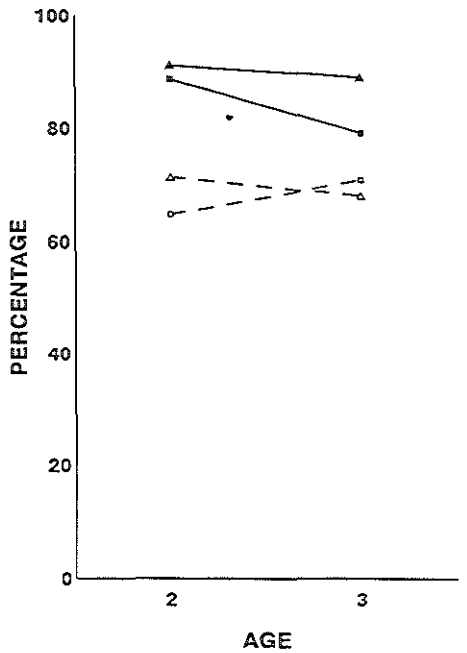
94. WAKES OFTEN



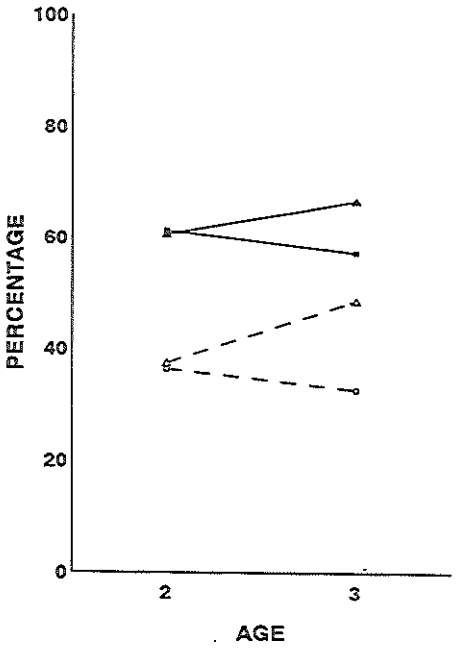
95. WANDERS AWAY FROM HOME



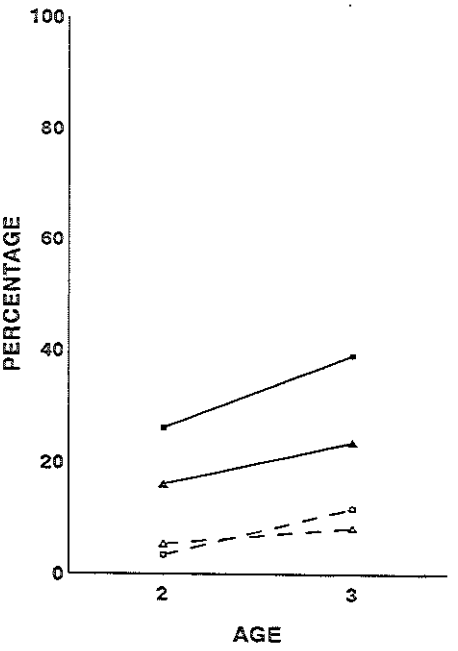
96. WANTS ATTENTION



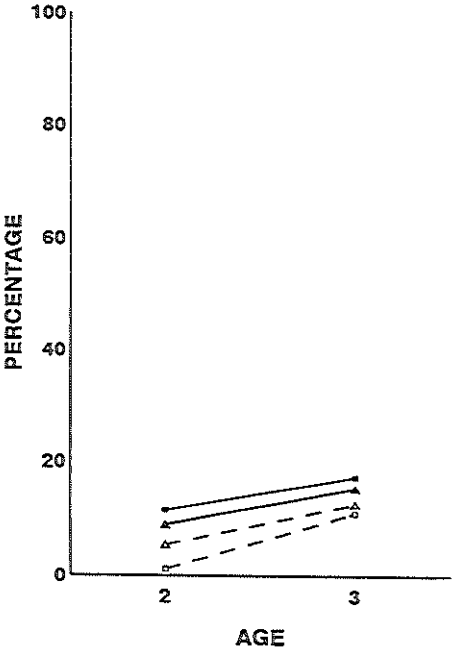
97. WHINING



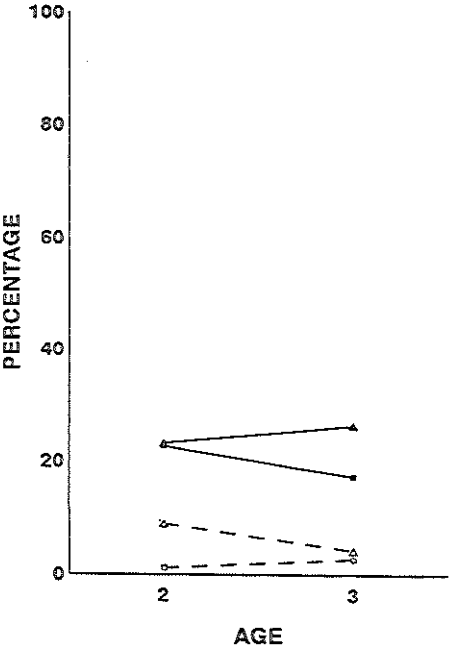
98. WITHDRAWN



99. WORRYING



100. OTHER PROBLEMS



Appendix H.1
Letter to the parents of children in the community sample - Time 2

Dear parent(s) / caregiver(s),

Maybe you remember that in 1989 we started a study among children ages 2-3 years in the province of Zuid-Holland. The goal of this study was to obtain information on the development of the behavior and competence of young children. In 1989 you kindly gave your cooperation to this study. Thanks to your participation we obtained data that appear to be very useful to health professionals. The information that you gave about your child was used to improve the mental health services provided to young children and their parents.

We now want to study which changes occur when the children enter the first years of elementary school. For this study we again need the help of the same parents and children, who participated in 1989. Therefore, we kindly ask you to give us one hour of your time. If you agree, one of our field-workers will visit you and ask you questions. Possibly she will also ask your child to perform a task together with her [1]. If possible, the field-worker will contact you by telephone during one of the months October through March. Like the first time, the information that you give will be recorded and used completely anonymously.

We sincerely hope that you will participate again in this important study. Doing this you will give an important contribution to the improvement of help given to children who have problems.

With kind regards,
yours sincerely,

J.M. Koot, psychologist

[1] Data on formal cognitive tests obtained from 200 children in this study were not reported in this thesis.

Appendix H.2
Letter to the teachers of children in the community sample - Time 2

Dear teacher,

In 1989 the Department of Child and Adolescent Psychiatry of the Sophia Children's Hospital / Erasmus University Rotterdam started a study among children aged 2-3 years in the province of Zuid-Holland. The goal of this study is to obtain information on the development of the behavior and competence of young children across the first six years of life. The information yielded by the study to this moment appears to be very useful to improve mental health services to young children. Children who participated in this study in 1989 were selected completely at random. We now want to track changes in the development in these children after entering the first years of elementary school.

The parents of one of the children in your class participate in this study. They gave permission to ask for your cooperation (see enclosed permission form). Apart from the parents, who were recently interviewed, the teacher is an indispensable source of information regarding the broad functioning of the child. Information that you may provide us is very important for the study.

Therefore, we kindly ask you some of your time to fill out the enclosed questionnaires about children of elementary school age. Please return the completed questionnaires using the enclosed envelope.

Some questions in one of the checklists may seem odd or superfluous to you. However, most of these questions regard children of other age-groups. Further, you may have insufficient information about the child. In that case it may be useful to wait some time before you fill out the questionnaire, in order to observe the child somewhat more.

Your participation is very important, because only when most all or most of the teachers of the children in this study particpate, the results will contribute to the improvement of services provided to children. If you have any questions please call us on phone number: ...

Yours sincerely,

J.M. Koot, psychologist

Appendix H.3
Proportions of children with deviant CBCL/2-3 scale scores at age 2-3 yrs and deviant CBCL/4-18 scale scores at age 4-5 yrs

		CBCL/4-18 scale												
CBCL/2-3 scale	n	Wth (17)	Som (18)	Anx (18)	Soc (22)	Tht (9)	Att (19)	Del (14)	Agg (20)	Int (55)	Ext (57)	TP (55)	I/E (88)	Any (77)
Oppositional	20	15 18	0 0	20 22	25 23	5 11	15 16	15 21	40 40	50 18	60 21	70 26	80 18	60 16
Withdrawn/Depressed	15	27 24	7 6	20 17	33 23	13 22	20 16	7 7	20 15	53 15	40 11	60 16	67 11	67 13
Aggressive	17	6 6	18 17	12 11	18 14	12 22	6 5	12 14	24 20	29 9	35 11	47 15	53 10	53 12
Anxious	16	19 18	6 6	13 11	13 9	6 11	13 11	6 7	0 0	44 13	13 4	31 9	56 10	44 9
Overactive	16	6 6	0 0	13 11	6 5	0 0	6 5	13 14	19 15	25 7	44 12	38 11	56 10	25 5
Sleep Problems	18	0 0	6 6	6 6	6 5	11 22	11 11	0 0	6 5	17 6	6 2	22 7	17 3	17 4
Somatic Problems	21	5 6	14 17	5 6	10 9	0 0	5 5	5 7	5 5	14 6	14 5	14 6	24 6	19 5
Internalizing	71	17 71	3 11	10 39	11 36	4 33	10 37	9 43	11 40	31 40	23 28	30 38	37 30	32 30
Externalizing	62	10 35	8 28	15 50	11 32	5 33	13 42	8 36	23 70	36 40	40 44	44 49	57 40	42 34
Total Problems	61	13 47	7 22	16 56	15 41	8 56	15 47	12 50	21 65	39 44	38 40	49 55	57 40	44 35
Internalizing and/or Externalizing	112	11 71	5 33	9 56	9 46	4 44	10 58	7 57	14 80	29 58	29 56	31 64	42 53	33 48
Any one or more	85	8 41	7 33	9 44	11 41	5 44	8 37	6 36	14 60	26 40	28 42	33 51	44 42	35 39

Note. N = 397. Wth = Withdrawn; Som = Somatic Complaints; Anx = Anxious/Depressed; Soc = Social Problems; Tht = Thought Problems; Att = Attention Problems; Del = Delinquent Behavior; Agg = Aggressive Behavior; Int = Internalizing; Ext = Externalizing; TP = Total Problems; I/E = Internalizing and/or Externalizing; Any = Any one or more. Numbers of children having deviant CBCL/4-18 scores are given in parentheses. The first row shows the proportion of children having deviant CBCL/2-3 scores also having deviant CBCL/4-18 scores; the second row shows the proportion of children having deviant CBCL/4-18 scores also having deviant CBCL/2-3 scores.
[a] Borderline cutpoint: T-score ≥ 67 for syndrome scales; T-score ≥ 60 for Internalizing, Externalizing, and Total problems.

Appendix H.4
Proportions of children with deviant CBCL/2-3 scale scores at age 2-3 yrs and deviant TRF scale scores at age 4-5 yrs

		TRF scale												
CBCL/2-3 scale	n	Wth (19)	Som (12)	Anx (15)	Soc (15)	Tht (15)	Att (17)	Del (15)	Agg (17)	Int (52)	Ext (52)	TP (53)	I/E (92)	Any (73)
Oppositional	20	5	5	0	10	5	10	5	15	20	30	30	40	25
		5	8	0	13	7	11	7	18	8	12	11	9	7
Withdrawn/Depressed	14	14	14	0	14	14	21	14	21	43	29	43	57	50
		11	17	0	13	13	18	13	18	12	8	11	9	10
Aggressive	15	0	0	13	13	13	0	13	13	33	27	20	40	27
		0	0	13	13	13	0	13	12	10	8	6	7	6
Anxious	16	13	6	6	0	13	6	6	6	25	13	19	38	31
		11	8	7	0	13	6	7	6	8	4	6	7	7
Overactive	16	6	0	19	0	0	19	6	6	25	25	31	38	31
		5	0	20	0	0	18	7	6	8	8	9	7	7
Sleep Problems	16	0	6	6	13	19	13	6	6	13	31	19	31	31
		0	8	7	13	20	12	7	6	4	10	6	5	7
Somatic Problems	19	0	0	5	5	5	5	0	0	16	16	11	32	16
		0	0	7	7	7	6	0	0	6	6	4	7	4
Internalizing	65	9	5	3	6	6	14	8	9	19	14	22	28	31
		32	25	13	27	27	53	33	35	23	17	26	20	27
Externalizing	60	8	2	8	7	7	8	8	12	22	27	25	38	28
		26	8	33	27	27	29	33	41	25	31	28	25	23
Total Problems	56	5	4	9	7	7	9	7	9	25	21	25	36	29
		16	17	33	27	27	29	27	29	27	23	26	22	22
Internalizing and/or Externalizing	104	8	3	6	5	7	11	7	9	19	19	22	33	29
		42	25	47	33	47	71	47	53	39	39	43	37	41
Any one or more	80	5	4	9	5	8	9	4	6	21	19	21	33	26
		21	25	47	27	40	41	23	29	33	29	32	28	29

Note. *N* = 349. Wth = Withdrawn; Som = Somatic Complaints; Anx = Anxious/Depressed; Soc = Social Problems; Tht = Thought Problems; Att = Attention Problems; Del = Delinquent Behavior; Agg = Aggressive Behavior; Int = Internalizing; Ext = Externalizing; TP = Total Problems; I/E = Internalizing and/or Externalizing; Any = Any one or more. Numbers of children having deviant CBCL/4-18 scores are given in parentheses. The first row shows the proportion of children having deviant CBCL/2-3 scores also having deviant CBCL/4-18 scores; the second row shows the proportion of children having deviant CBCL/4-18 scores also having deviant CBCL/2-3 scores.
[a] Borderline cutpoint: T-score ≥ 67 for syndrome scales; T-score ≥ 60 for Internalizing, Externalizing, and Total problems.

Appendix H.5
Correlations among predictors

	DIS	GD	OPP	WIT	AGG	ANX	OVE	TP	LE	CH	TEMP	LP	SC
Disadvantage	1.00												
General development	.03	1.00											
Oppositional	.41	-.05	1.00										
Withdrawn/Depressed	.24	-.17	.36	1.00									
Aggressive	.33	-.01	.49	.28	1.00								
Anxious	.19	-.14	.35	.43	.12	1.00							
Overactive	.28	-.13	.53	.30	.40	.19	1.00						
Total Problems	.45	-.11	.85	.57	.62	.55	.64	1.00					
Life-events	.14	.04	.14	.11	.09	.07	.15	.20	1.00				
Child health	.20	-.03	.11	.10	.03	.11	.13	.14	.09	1.00			
Temperament	.14	-.12	.23	.17	.16	.22	.32	.31	.18	.15	1.00		
Language problems	.09	-.23	.14	.23	.17	.19	.19	.25	.08	.14	.27	1.00	
School competence	-.11	.16	-.13	-.20	-.11	-.12	-.17	-.20	.09	-.16	-.15	-.46	1.00
Parenting stress	.21	-.18	.34	.17	.31	.14	.35	.37	.15	.14	.46	.26	-.21

Note: N = 307; DIS = Disadvantage; GD = General development; OPP = Oppositional; WIT = Withdrawn/Depressed; AGG = Aggressive; ANX = Anxious; OVE = Overactive; TP = Total Problems; LE = Life-events; CH = Child health; TEMP = Temperament; LP = Language problems; SC = School competence.
Significance levels of r: $p < .01$ with $r = .12$; $p < .001$ with $r = .16$.

Appendix H.6
Correlations between predictors (excluding Time 1 CBCL/2-3 scales) and Time 2 CBCL/4-18 and TRF scales

	CBCL/4-18 scales										
	WTH	SOM	AXD	SOC	THT	ADD	DEL	AGG	INT	EXT	TP
Disadvantage	.18	.11	.22	.21	.08	.24	.18	.27	.24	.27	.30
General development	-.04	.00	.03	-.16	.00	-.13	.01	.02	.00	.02	-.01
Life-events	.17	.06	.22	.16	.14	.26	.20	.25	.22	.26	.32
Child Health	.21	.25	.23	.22	-.01	.21	.08	.19	.30	.18	.25
Temperament	.26	.11	.25	.27	.18	.39	.20	.38	.29	.37	.43
Language problems	.24	.03	.10	.38	.14	.38	.23	.20	.16	.22	.30
School competence	-.14	-.10	-.13	-.36	-.19	-.35	-.11	-.20	-.16	-.19	-.27
Parenting stress	.25	.14	.30	.41	.20	.53	.32	.46	.33	.47	.53
	TRF scales										
	WTH	SOM	AXD	SOC	THT	ADD	DEL	AGG	INT	EXT	TP
Disadvantage	.10	.09	.07	.03	.05	.12	.09	.05	.11	.06	.11
General development	-.07	-.06	-.14	-.12	-.18	-.01	-.05	-.09	-.05	-.05	-.13
Life-events	.03	-.00	.15	.16	.18	.19	.03	.11	.11	.10	.17
Child health	.10	.14	.05	.14	.07	.09	.06	.05	.10	.06	.10
Temperament	.09	-.01	.10	.15	.17	.17	.05	.07	.11	.07	.15
Language problems	.30	.00	.09	.33	.13	.46	.07	.08	.20	.08	.30
School competence	-.44	-.15	-.38	-.56	-.25	-.64	-.23	-.42	-.48	-.42	-.61
Parenting stress	.07	.01	.08	.21	.14	.22	.20	.23	.08	.24	.25

Note: $N = 348$.
WTH = Withdrawn; SOM = Somatic Complaints; AXD = Anxious/Depressed; SOC = Social problems; THT = Thought problems;
ADD = Attention problems; DEL = Delinquent behavior; AGG = Aggressive behavior; INT = Internalizing;
EXT = Externalizing; TP = Total problems.
Significance levels of r: $p < .01$ with $r = .13$; $p < .001$ with $r = .17$.

Summary

The objective of the studies reported in this thesis was to assess the value of a standardized parent rating scale for behavioral and emotional problems in young preschoolers and to study the prevalence and correlates of problem behavior in children 2- to 3-years old. It was argued that it is difficult to diagnose problem behavior in young preschoolers as deviant or disordered due to the fluctuations and relative commonness of difficult behavior at this age, the dependency of its description on diverse informants, and the lack of a validated taxonomy of such behavior. In chapter 2 an overview was given of efforts to standardize the assessment of problem behavior in preschoolers, and the results of relevant studies on its prevalence and longitudinal course were discussed. The main limitations of existing parent and teacher rating scales of preschoolers' behavioral/emotional problems appeared to be that they do not provide a well differentiated picture of particular problem areas, and that no data on their validity for very young children were reported. In the Netherlands, no validated instrument to assess problem behavior in 2-3-year-olds was available. In recent years, however, a 99-item parent rating scale, the Child Behavior Checklist for Ages 2-3 (CBCL/2-3) was developed in the United States, on which problem behavior can be rated on a three-point (0, 1, 2) scale. The rating scale yielded six different problem behavior syndromes (Achenbach, Edelbrock, & Howell, 1987; Achenbach, 1992). The studies in this thesis assessed the usefulness of this instrument in the Dutch situation. It was deemed necessary to establish the cross-national comparability of the CBCL/2-3 syndromes; to assess the psychometric properties of its syndrome scales in a Dutch population, i.e., the factor structure of the instrument, internal consistency, test-retest stability, and interrater agreement; construct and criterion related validity, and predictive validity. For diagnostic and clinical applications, and for some research applications, Dutch normative data on the instrument are needed. Further, two studies in this thesis reported on the prevalence and correlates of problem behavior in different samples using the CBCL/2-3.

The samples and methods of data collection used in the studies were described in chapter 3. Data from three samples were used: a representative sample of 420 children (215 boys, 205 girls) from the general population of the Dutch province of Zuid-Holland (the community sample); a sample of 458 children (302 boys, 156 girls) referred to 12 different mental health settings (child guidance, community mental health, child psychiatry, and medical day-care; the clinical sample); and a sample of 1306 twin pairs (1291 boys, 1321 girls).

The study reported in chapters 4 and 5 had the objective to obtain psychometric data

on the Dutch translation of the CBCL/2-3. This was accomplished by studying its factor structure in the three samples; by studying its internal consistency and test-retest reliability; by studying its relation to other instruments designed to measure a similar construct (convergent measure) and dissimilar phenomena (divergent measure); and by studying its power to differentiate between normal children and children with acknowledged behavioral/emotional problems. Exploratory and confirmatory factor analyses yielded a factor structure that was highly congruent across different samples. The analyses supported the differentiation of young children's problem behaviors into six syndromes reflecting oppositional behavior, aggression, and overactivity, constituting an externalizing dimension, and anxiety and withdrawal, constituting an internalizing dimension. Sleep problems represent a separate construct at this age. An expected somatic problems syndrome was not replicated. Scores computed for children from the community sample on American and Dutch CBCL/2-3 syndrome scales showed high cross-cultural similarity of scale scores. The internal consistency and test-retest reliability of all scales was good, except for withdrawn/depressed behavior, where it was moderate. The convergent construct validity was supported by significant associations between the CBCL/2-3 and the English Behaviour Checklist (BCL; Richman, 1977) using both continuous and categorical scores. The mean correlation between the BCL and the total problem score was .65. Classification of the 3-year-olds in the community sample according to the appropriate clinical cutoff points showed an 86.7 percent agreement across instruments. The divergent construct validity of the instrument was confirmed by low negative correlations with a measure for general development, the Minnesota Child Development Inventory. The criterion-related validity of the CBCL/2-3 was supported by the significant differences between demographically matched referred and nonreferred children's scale scores. Regression analyses showed medium effects of referral status for both sexes. Small but significant effects of socioeconomic status (SES) were found, indicating higher scores for boys and girls from lower SES on most scales. Small but significant effects of age indicated higher scores for older children on some scales, most notably among boys.

The studies reported in chapters 6 an 7 regarded the prevalence of problem behavior in Dutch toddlers aged 2-3 years in the community and clinical sample as measured by the CBCL/2-3. Further, it was investigated how CBCL/2-3 scores in both samples related to demographic, family, and child characteristics, and in the clinical sample also to clinically relevant variables, such as referral complaints and psychiatric diagnosis. Both studies provided data that until now were not available for the Netherlands.

In the community sample, certain behaviors commonly regarded as problematic were reported by parents to be present in more than fifty percent of the preschool children. These items may thus be regarded as typical for the age group studied. They reflect the limited emotional and behavioral control, and resistance to socialization, which may be expected in the typical young preschooler. Because so many items were scored 1 or 2, the presence of those behaviors as such may not indicate a behavioral/emotional problem. For only 18 items, however, 10 percent or more of the children received score 2. For most items a score 2 may thus more validly indicate that the child's behavior may be regarded

a serious problem. Some items were scored by so few parents, that the children for whom this behavior was reported may be in need of special attention. Most of these items reflect withdrawn, depressed, and aggressive behaviors, and somatic complaints. In this sample, 12.4% percent of the 2-3-year olds had one or more syndrome scores above the clinical cutpoint (98th percentile), and 21.6% above the borderline cutpoint (95th percentile). A robust finding was that boys showed more aggressive behaviors, and 2-3-year olds from lower SES strata showed more problem behaviors of several kind than children from higher strata.

In the community sample, parental education, socioeconomic position of the family, health of the child, parental mental health, and maternal parenting attitude were the strongest correlates of young preschoolers' problem behavior as assessed by the CBCL/2-3. Factors were selectively related to syndrome scores. Internalizing problems were specifically related to low parental education, ill health of the child, poor physical health of the mother, and poor mental health in both parents. Externalizing behaviors were related to low educational and socioeconomic status of the family, ill health of the child, mother being exhausted, irritated, and physical punishing the child, and the presence of stressful life-events. Sleep problems were specifically related to ill health of the child, number of days in hospital, poor maternal health, and marital conflicts. Somatic problems held a specific relation to maternal physical and mental health.

In the sample of clinically referred children, mean CBCL/2-3 scores were on the average almost two times as high as in the community sample. Small sex, age, and SES effects on scale scores found in this sample were barely comparable to those obtained in the community sample. Almost 70% of the children had one or more deviant syndrome scores using borderline range cutpoints. Unlike in the community sample, deviant scale scores concurred with deviant total problem scores in at least 80% of all cases. Using the same cutpoint for both sexes, girls had a somewhat higher proportion of deviant sleep problems and somatic problems scores. However, the large difference between boys and girls on proportion of deviant aggressive scores was not present in this sample.

Comparison of the findings on correlates of problem behavior in the clinical and community sample showed that indices of educational/occupational level of the parents, indices of child health or hospitalization, parental mental health, caregiving, and single parenting were related to the same problem scales in both samples. Of the narrow-band externalizing scales the oppositional scale was related to (quality) of caregiving, aggressive behavior to parental education, child health and parental mental health, and overactive behavior to parental education, child health and (quality) of care in both samples. Of the narrow-band internalizing scales both withdrawn and anxious behavior were only consistently related to indices of poor health or hospitalization. The only correlate of sleep problems recurring in both samples was single parenthood. For somatic problems no recurring correlates were found. Consistent correlates of the broad-band Internalizing scale were parental education, child health and parental mental health. Both Externalizing and total problem scores showed consistent relations to parental education, child health and caregiving history.

Most children in the clinical sample were referred to mental health services by primary and specialist health care. The majority of the sample presented with conduct problems (56.1%), developmental problems (34.7%), and/or management problems (27.3%). Based on the clinical file information present at the services almost 40% of the children received a psychiatric diagnosis, the majority of which were hyperkinetic (23%), conduct (30%) and emotional (20%) disorders.

Problem behaviors as reported by the parents were significantly related to type of mental health service. For children referred to semi-residential settings higher levels of internalizing problems were reported than for children referred to other services. The children referred to the daycare centers and the child guidance consultation agency had more externalizing problems than those referred to community mental health and child psychiatry units. The four types of services differed in predictable ways with regard to the referral source of the clients, the number of services previously received by the clients, the type of referral question, and the type of referral complaints. Therefore, the differences between types and levels of problem behavior among the children referred to the various services may be regarded as evidence of the discriminant validity of CBCL/2-3 syndrome scores. CBCL/2-3 scales were also differentially and in predictable directions related to referral complaints and ICD-10 diagnoses based on file information. This is an important finding, because it relates parent reports on problem behavior to information on problems obtained from independent sources.

Chapter 8 regarded the stability and predictive validity of toddlers' problem behavior as measured by the CBCL/2-3. This was done in a prospective follow-up study of the community sample two years after initial assessment. The stability and change of parents' reports of children's problem behavior was assessed over the two-year period. Further, it was determined how well parents' initial reports of problem behavior (Time 1) predicted parents' and teachers' reports of problem behavior and signs of significant problems (Time 2), over and above family and child characteristics and intervening stressful experiences. Time 1 parent ratings of CBCL/2-3 scales significantly predicted Time 2 parent ratings of CBCL/4-18 scales. All Time 1 scales except those for sleep problems and somatic problems accounted for medium amounts of variance in Time 2 scale scores. Cross-time correlations of Externalizing and total problem scores were large, and showed significantly larger stabilities for these scores than for Internalizing scores. These findings were supported by categorical analyses. Of the children who had deviant CBCL/2-3 Internalizing, Externalizing, or total problem scores at Time 1, the percentages who had a deviant score on the corresponding Time 2 parent rating scale were 31%, 40%, and 49%, respectively. When used as a screening test, the CBCL/2-3 total problem score showed a predictive sensitivity of 63%, and a predictive specificity of 84%, for an overall correct prediction of 73%.

Predictive correlations between Time 1 parent ratings and Time 2 teacher ratings were low. However, they showed an overall pattern similar to that found for stabilities of parent ratings. Predictive relations were lower for Internalizing than for Externalizing behavior. Categorical analyses showed that deviant Time 1 parent ratings predicted

deviant Time 2 teacher ratings in only 21%-27% of the cases using Internalizing, Externalizing, or total problem scores.

As was expected, Time 2 parent reports of health problems, difficult temperament and parenting stress, and language problems and teacher rated school competence were all related to Time 2 parent ratings of behavioral/emotional problems. However, despite medium to high concurrent correlations with the Time 2 syndromes, in only two instances these predictors were among the strongest for a given syndrome. For seven out of eight Time 2 parent ratings of syndromes, the Time 1 CBCL/2-3 scales were the strongest predictor over and above family characteristics and other child characteristics, even if these were measured concurrent to the Time 2 syndromes. Over and above teacher rated school competence, the CBCL/2-3 scales for aggressive and overactive behavior had an independent contribution to some teacher rated behavior syndromes. A remarkable finding was that parent reported stressful life-experiences had a direct contribution to the development of behavioral/emotional problems as rated by both parents and teachers, especially of an anxious/depressed and attentional nature.

The longitudinal study provided suggestive evidence that those 2-3-year-olds who initially show problems at a high level, comparable to that of children referred for mental health and child psychiatric services, and who are more likely to come from lower SES groups, have a difficult temperament, poor health, and to experience more stressful life-events may be especially at risk for having behavioral/emotional problems at age 4-5.

In chapter 9 conclusions and implications of the study's results were discussed. It was concluded that the CBCL/2-3 can be used as a broad-band instrument that gives a valid first impression of the type and severity of psychopathology in the 2-3-year-old child, especially immediately after referral and during evaluation of treatment. Beside the CBCL/2-3 other, more specific instruments may be used to cover more specific psychopathology (e.g., autism), and to cover other domains of functioning. The CBCL/2-3 covers several aspects of deviancy that can be compared to sex-appropriate norms: frequency, intensity, and constellation of behavioral/emotional problems. Other parameters of deviancy such as chronicity of the problems and the (social) context in which they occur are not covered by the instrument, but should be explored through observations and direct contact with diverse informants.

Samenvatting

Het doel van het onderzoek, waarover in dit proefschrift gerapporteerd werd, was om de waarde vast te stellen van een gestandaardiseerde oudervragenlijst voor de beoordeling van emotionele en gedragsproblemen en om het vóórkomen van probleemgedrag bij kinderen van 2-3 jaar en daarmee samenhangende variabelen te onderzoeken. Het is moeilijk is om probleemgedrag bij peuters en jonge kleuters te diagnosticeren als afwijkend of gestoord, vanwege de fluctuaties in moeilijk gedrag op deze leeftijd en vanwege het feit dat het om relatief gewoon en veel voorkomend gedrag gaat. Voorts is de beschrijving van dit gedrag afhankelijk van verschillende informanten en ontbreekt een gevalideerd ordeningssysteem voor zulk gedrag. In hoofdstuk 2 werd een overzicht gegeven van gestandaardiseerde methoden ontwikkeld voor de vaststelling van probleemgedrag bij peuters en jonge kleuters, en de resultaten van relevante onderzoeken over het vóórkomen en het longitudinale beloop ervan werden besproken. De belangrijkste beperkingen van schalen voor de beoordeling van emotionele en gedragsproblemen bij kleuters door ouders en leerkrachten bleken te zijn dat ze geen goed gedifferentieerd beeld van de verschillende probleemgebieden geven, en dat er geen gegevens zijn gerapporteerd over hun geschiktheid voor gebruik bij zeer jonge kinderen. In Nederland was helemaal geen gevalideerd instrument beschikbaar om probleemgedrag bij 2-3-jarigen vast te stellen. Recent werd echter in de Verenigde Staten een schaal voor de beoordeling van kindgedrag door ouders ontwikkeld, de Child Behavior Checklist for Ages 2-3 (CBCL/2-3), bestaande uit 99 items, die beoordeeld worden met 0 (helemaal niet van toepassing), 1 (een beetje of soms van toepassing), of 2 (duidelijk of vaak van toepassing). Deze schaal onderscheidt zes verschillende syndromen voor probleemgedrag (Achenbach, 1992; Achenbach, Edelbrock, & Howell, 1987). Door middel van de onderzoeken in dit proefschrift werd de bruikbaarheid van dit instrument in de Nederlandse situatie onderzocht. Daartoe werd nagegaan hoe vergelijkbaar de CBCL/2-3 syndromen in de twee landen zijn en welke de psychometrische eigenschappen van de syndroomschalen in een Nederlandse populatie zijn, dat wil zeggen, de factor structuur van het instrument, de interne consistentie, de test-hertest stabiliteit, en de overeenstemming tussen beoordelaars; de constructvaliditeit, de criteriumgerelateerde, en de predictieve validiteit. Nederlandse normgegevens, nodig voor diagnostische en klinische toepassingen, en voor sommige toepassingen in onderzoek, werden voor het instrument verzameld. Voorts hadden twee onderzoeken in dit proefschrift betrekking op het vóórkomen van probleemgedrag in verschillende groepen jonge kinderen, zoals vastgesteld met de CBCL/2-3, en de samenhang daarvan met een aantal

achtergrondgegevens.

De steekproeven en de in het onderzoek gebruikte methoden van gegevensverzameling werden beschreven in hoofdstuk 3. Er werden gegevens uit drie steekproeven gebruikt: een representatieve steekproef van 420 kinderen (215 jongens, 205 meisjes) uit de algemene bevolking van de provincie Zuid-Holland (de bevolkingssteekproef); een steekproef van 458 kinderen (302 jongens, 156 meisjes) verwezen naar 12 verschillende instellingen voor de geestelijke gezondheidszorg (pedagogisch consultatiebureau, riagg's, afdelingen kinderpsychiatrie, en medisch kleuterdagverblijven; de klinische steekproef); en een steekproef bestaande uit 1306 tweelingparen (1291 jongens, 1321 meisjes).

Het onderzoek dat gerapporteerd werd in de hoofdstukken 4 en 5 had de verzameling van psychometrische gegevens over de Nederlandse vertaling van de CBCL/2-3 als doel. We onderzochten de factorstructuur van de lijst in de drie steekproeven; de interne consistentie en test-hertest betrouwbaarheid van de vragenlijst; de relatie ervan met andere instrumenten, ontworpen om een soortgelijk construct (convergente maat) en ongelijksoortige construct (divergente maat) te meten; en inhoeverre het instrument onderscheid kan aanbrengen tussen normale kinderen en kinderen met geïdentificeerde emotionele en gedragsproblemen. Exploratieve en bevestigende factoranalyses leverden een factorstructuur op die sterk overeenkwam in de verschillende steekproeven. De analyses ondersteunden de differentiatie van probleemgedrag van jonge kinderen in zes syndromen, die te maken hebben met opstandig gedrag, agressie, en hyperactiviteit, welke samen een externaliserende dimensie vormen, en angst en teruggetrokken gedrag, welke samen een internaliserende dimensie vormen. Slaapproblemen vertegenwoordigen op deze leeftijd een afzonderlijk construct. Tegen de verwachting in werd geen apart syndroom voor lichamelijke problemen gevonden. Scores berekend voor kinderen uit de bevolkingssteekproef op de Amerikaanse en Nederlandse syndroomschalen toonden een sterke cross-culturele gelijkenis van de schalen. De interne consistentie en test-hertest betrouwbaarheid van alle schalen was goed, behalve voor die voor teruggetrokken/depressief gedrag, waarvoor deze matig waren. De convergente constructvaliditeit werd ondersteund door een samenhang tussen de CBCL/2-3 en de Engelse Behaviour Checklist (BCL; Richman, 1977), zowel op het niveau van continue als categorale scores. De gemiddelde correlatie tussen de BCL en de CBCL totale probleemscore was .65. Klassifikatie van de driejarigen in de bevolkingssteekproef volgens de geijkte klinische afsnijdpunten liet een overeenstemmingspercentage van 86,7 tussen beide instrumenten zien. De divergente constructvaliditeit van het instrument werd bevestigd door lage negatieve correlaties met een maat voor algemene ontwikkeling, de Minnesota Child Development Inventory. De criterium-gerelateerde validiteit van de CBCL/2-3 werd ondersteund door de significante verschillen tussen de schaalscores van op demografische variabelen gematchte verwezen en niet-verwezen kinderen. Regressieanalyses toonden middelmatige effecten van verwezen-zijn voor beide geslachten. Kleine maar significante effecten van sociaal-economisch niveau gaven op de meeste schalen hogere scores voor jongens en meisjes uit lagere sociaal-economische strata te zien. Kleine maar significante effecten van leeftijd duidden op hogere scores voor oudere kinderen op sommige schalen, vooral onder jongens.

De onderzoeken gerapporteerd in de hoofdstukken 6 en 7 hadden betrekking op probleemgedrag bij Nederlandse 2-3-jarige peuters in de algemene bevolkings- en klinische steekproeven, zoals gemeten met de CBCL/2-3. Voorts werd onderzocht hoe CBCL/2-3 scores in beide steekproeven samenhingen met demografische, gezins- en kindkenmerken, en in de klinische steekproef ook met klinisch relevante variabelen, zoals verwijzingsklachten en psychiatrische diagnose. Beide onderzoeken leverden gegevens op waarover we in Nederland tot nu toe niet beschikten.

In de bevolkingssteekproef werden sommige gedragingen, die men gewoonlijk problematisch vindt, door ouders gerapporteerd voor meer dan vijftig procent van de peuters. Deze gedragingen kunnen dus beschouwd worden als typisch voor de onderzochte leeftijdsgroep. Ze weerspiegelen de beperkte emotionele en gedragsmatige controle en de weerstand tegen socialisatie, die je bij de typische peuter kunt verwachten. Omdat zoveel items met 1 of 2 gescoord werden duidt de aanwezigheid van dergelijke gedragingen als zodanig niet op een emotioneel of gedragsprobleem. Echter, bij slechts 18 items kregen 10 procent of meer van de kinderen een score 2. Bij de meeste items is een score 2 wellicht dus een meer valide indicatie dat het gedrag van het kind beschouwd kan worden als een ernstig probleem. Sommige items werden door zo weinig ouders gescoord dat kinderen voor wie dit gedrag werd gerapporteerd wellicht speciale aandacht nodig hebben. De meeste van deze items weerspiegelen sterk teruggetrokken, depressieve en agressieve gedragingen, en lichamelijke klachten. In deze steekproef hadden 12,4% van de 2-3-jarigen een of meer syndroomscores boven het klinische afsnijdpunt (98ste percentiel) en 21,6% boven het afsnijdpunt voor het grensgebied (95ste percentiel). Een sterke bevinding was dat jongens meer agressieve gedragingen vertoonden, en dat 2-3-jarigen uit bevolkingsgroepen van lager sociaal-economisch niveau meer probleemgedragingen van allerlei aard lieten zien dan andere kinderen.

In de bevolkingssteekproef waren opleiding van de ouders, sociaal-economische positie van het gezin, gezondheid van het kind, geestelijke gezondheid van de ouders en opvoedingshouding van de moeder het sterkst gerelateerd aan probleemgedrag van peuters, zoals vastgesteld met de CBCL/2-3. Deze factoren hadden een specifieke relatie met syndroomscores. Internaliserende problemen hadden een specifieke relatie met lage opleiding van de ouders, slechte gezondheid van het kind, slechte lichamelijke gezondheid van de moeder, en slechte geestelijke gezondheid van beide ouders. Externaliserende gedragingen waren gerelateerd aan een laag opleidings- en sociaal-economisch niveau van het gezin, slechte gezondheid van het kind, een moeder die uitgeput en geïrriteerd is, en het kind lichamelijk straft, en de aanwezigheid van stressvolle gebeurtenissen. Slaapproblemen waren specifiek gerelateerd aan slechte gezondheid van het kind, aantal dagen doorgebracht in een ziekenhuis, slechte gezondheid van de moeder, en huwelijksconflicten. Lichamelijk problemen hadden een specifiek verband met de lichamelijke en geestelijke gezondheid van de moeder.

In de steekproef van naar GGZ-instellingen verwezen kinderen waren de gemiddelde CBCL/2-3 scores bijna tweemaal zo hoog als in de bevolkingssteekproef. De kleine verschillen in scores toe te schrijven aan geslacht, leeftijd, en sociaal-economische positie

waren nauwelijks vergelijkbaar met die in de algemene bevolking. Bijna 70% van de kinderen had een of meer afwijkende syndroomscores volgens het grensgebied-criterium. Anders dan in de bevolkingssteekproef vielen afwijkende schaalscores in tenminste 80% van alle gevallen samen met een afwijkende totale probleemscore. Met hetzelfde afsnijdpunt voor beide geslachten als criterium hadden meisjes een wat grotere proportie afwijkende scores bij slaapproblemen en lichamelijke klachten dan jongens. Het grote verschil tussen jongens en meisjes wat betreft het percentage afwijkende scores op de 'aggressief' schaal werd in deze steekproef echter niet gevonden.

Vergelijking van de bevindingen over aan probleemgedrag gerelateerde variabelen in de klinische en bevolkingssteekproef liet zien dat indices voor het opleidings- en beroepsniveau van de ouders, indices voor de gezondheid van het kind of opname in een ziekenhuis, geestelijke gezondheid van de ouder, stabiliteit en kwaliteit van de zorg, en het eenouderschap in beide steekproeven aan dezelfde probleemschalen waren gelieerd. Van de specifieke schalen voor externaliserend gedrag was in beide steekproeven de schaal voor oppositioneel gedrag gerelateerd aan de kwaliteit van de zorg, die voor agressief gedrag aan de opleiding van de ouders, gezondheid van het kind, en geestelijke gezondheid van de ouders, en die voor hyperactiviteit aan opleiding van de ouders, gezondheid van het kind en kwaliteit van de zorg. Van de specifieke schalen voor internaliserend gedrag waren zowel teruggetrokken als angstig gedrag alleen consistent gerelateerd aan indices voor slechte gezondheid of opname in een ziekenhuis. De enige factor die in beide steekproeven gerelateerd was aan slaapproblemen was het eenouderschap. Er werden geen factoren gevonden die in beide steekproeven gerelateerd waren aan lichamelijke klachten. Terugkerende factoren gerelateerd aan de schaal voor internaliserend gedrag waren opleiding van de ouders, gezondheid van het kind en geestelijke gezondheid van de ouders. Zowel scores voor externaliserend gedrag als de totale probleemscore waren in beide steekproeven gelieerd aan opleiding van de ouders, gezondheid van het kind en stabiliteit of kwaliteit van de zorg.

De meeste kinderen in de klinische steekproef werden naar GGZ-instellingen verwezen door vertegenwoordigers van de eerstelijns en tweedelijns gezondheidszorg. De meerderheid van de steekproef werd aangemeld met gedragsproblemen (56.1%), ontwikkelingsproblemen (34.7%), en/of moeilijk hanteerbaar gedrag (27.3%). Gebaseerd op dossierinformatie aanwezig bij de instellingen kreeg bijna 40% van de kinderen een psychiatrische diagnose, waarvan de meerderheid hyperkinetische (23%), gedrags- (30%) en emotionele stoornissen (20%) waren.

Probleemgedrag gerapporteerd door ouders was significant gerelateerd aan het type GGZ-instelling. Voor kinderen verwezen naar semi-residentiële voorzieningen, de medische kleuterdagverblijven, werden hogere niveau's van internaliserende problemen gerapporteerd dan voor kinderen verwezen naar andere instellingen. De kinderen verwezen naar dagverblijven en naar het pedagogisch consultatiebureau hadden meer externaliserende problemen dan de kinderen die verwezen waren naar de riagg's en de afdelingen kinderpsychiatrie. De vier typen instellingen verschilden in voorspelbare richtingen met betrekking tot de bron van verwijzing, het aantal instellingen dat de cliënt eerder had

bezocht, het type verwijzingsvraag, en het type klachten bij verwijzing. Daarom kunnen de verschillen tussen typen en niveau's van probleemgedrag tussen kinderen verwezen naar de verschillende instellingen worden beschouwd als evidentie voor de discriminante validiteit van de CBCL/2-3 syndroomscores. De CBCL/2-3 schalen waren ook differentieel en in voorspelbare richtingen gerelateerd aan verwijzingsklachten en ICD-10 diagnoses gebaseerd op dossierinformatie. Hiermee is een verband aangetoond tussen het oordeel van ouders over probleemgedrag en informatie over problemen verkregen uit onafhankelijke bronnen.

Hoofdstuk 8 had betrekking op de stabiliteit en predictive validiteit van het probleemgedrag van peuters, zoals gemeten met de CBCL/2-3. Dit werd onderzocht in een prospectief vervolgonderzoek van de bevolkingssteekproef twee jaar na de beginmeting. De stabiliteit en verandering van de oordelen van ouders over probleemgedrag van kinderen werd vastgesteld over de periode van twee jaar. Voorts werd bepaald hoe goed de oordelen van ouders over probleemgedrag op Tijdstip 1 (T1) de oordelen van ouders en leerkrachten over probleemgedrag en tekenen van ernstige problemen op Tijdstip 2 (T2) voorspelden, onafhankelijk van gezins- en kindkenmerken en stressvolle ervaringen in de tussenliggende periode. Ouderoordelen op CBCL/2-3 schalen op T1 voorspelden ouderoordelen over probleemgedrag op T2, zoals gemeten met de schalen van de CBCL/4-18. Alle T1 schalen, behalve die voor slaapproblemen en lichamelijke klachten, verklaarden middelmatige proporties variantie in de T2 schaalscores. De correlaties over de tijd waren hoog voor externaliserend scores en totale probleemscores. Deze laatste toonden een sterkere stabiliteit dan internaliserend scores. Deze bevindingen werden ondersteund door categorale analyses. Van de kinderen met afwijkende CBCL/2-3 internaliserend, externaliserend en totale probleemscores hadden er respectievelijk 31%, 40% en 49% afwijkende scores op de overeenkomende ouderoordelen op T2. Bij gebruik als screeningstest met de T2 totale probleemscore als criterium werd voor de CBCL/2-3 totale probleemscore een sensitiviteit van 63%, een specificiteit van 84%, en totale juiste predictie van 73% gevonden.

De predictieve correlaties tussen ouderoordelen op T1 en leerkrachtoordelen op T2 waren laag. Ze gaven echter over het geheel een patroon te zien dat leek op dat gevonden voor de stabiliteit van ouderoordelen. De predictieve relaties waren lager voor internaliserend dan voor externaliserend gedrag. Categorale analyses toonden aan dat afwijkende leerkrachtoordelen op T2 voor internaliserende, externaliserende, en totale probleemscores in slechts 21%-27% van de gevallen voorspeld werden door afwijkende ouderoordelen op deze schalen op T1.

Zoals verwacht waren ouderoordelen over gezondheidsproblemen van het kind op T2, moeilijk temperament en opvoedingsbelasting, en taalproblemen en schoolvaardigheid zoals beoordeeld door de leerkracht alle gerelateerd aan ouderoordelen over emotionele en gedragsproblemen op T2. Ondanks hun middelmatige tot hoge gelijktijdige correlaties met de syndromen op T2 behoorden deze predictoren slechts in twee gevallen tot de sterkste voor een bepaald syndroom. Bij zeven van de acht ouderoordelen op syndromen op T2 waren de CBCL/2-3 schalen op T1 de sterkste predictor, onafhankelijk van gezinskenmer-

ken en andere kindkenmerken, zelfs als deze gelijktijdig met de syndromen op T2 werden gemeten. De CBCL/2-3 schalen voor agressief en hyperactief gedrag hadden een onafhankelijke bijdrage aan de voorspelling van gedragssyndromen beoordeeld door de leerkracht, ongeacht de schoolvaardigheid zoals beoordeeld door de leerkracht. Een opmerkelijke bevinding was dat door ouders gerapporteerde stressvolle gebeurtenissen een directe bijdrage hadden aan de ontwikkeling van emotionele en gedragsproblemen zoals beoordeeld door zowel ouders als leerkrachten, met name die van angstig/depressieve aard en aandachtsproblemen.

De resultaten uit het longitudinale onderzoek suggereerden dat naarmate het niveau van problemen op 2-3-jarige leeftijd hoger was (vergelijkbaar met dat van kinderen verwezen naar de GGZ en kinderpsychiatrische instellingen) en als kinderen uit lager sociaal-economisch milieu kwamen, een moeilijk temperament en een slechte gezondheid hadden, en meer stressvolle gebeurtenissen meegemaakt hadden, zij mogelijk een verhoogd risico liepen om op de leeftijd van 4-5 jaar emotionele en gedragproblemen te hebben.

In hoofdstuk 9 werden de conclusies en implicaties voortkomend uit de bevindingen van het onderzoek besproken. Geconcludeerd werd dat de CBCL/2-3 kan worden gebruikt als een instrument dat een valide eerste indruk geeft van het type en de ernst van psychopathologie bij het 2-3-jarige kind, vooral onmiddellijk na verwijzing en bij de evaluatie van behandeling. Behalve de CBCL/2-3 kunnen andere, meer specifieke instrumenten worden gebruikt, die specifiekere psychopathologie (zoals autisme) en andere gebieden van functioneren bestrijken. De CBCL/2-3 beslaat verschillende aspecten van afwijkend gedrag die met geslachtsspecifieke normen kunnen worden vergeleken: frequentie, intensiteit en samenstelling van emotionele en gedragsproblemen. Andere aanwijzingen voor de mate van afwijkendheid zoals de chroniciteit van de problemen en de (sociale) context waarin deze optreden worden niet door het instrument bestreken, maar moeten verder worden onderzocht door middel van observaties en in direct contact met verschillende informanten.

Dankwoord

Dit proefschrift is weliswaar geschreven door één persoon, maar dankt zijn ontstaan aan de bijdragen van velen. Ik ben blij hen daarvoor op deze plaats te kunnen bedanken.

Prof. Dr. Kees van Lieshout en Dr. Marianne Riksen-Walraven zetten mij al tijdens mijn studietijd op het pad van de wetenschap. Hun nooit aflatende steun in het eerste traject van mijn academische loopbaan gaf mij de onontbeerlijke "basic trust" als onderzoeker. Mede daaraan is het te danken dat dit proefschrift er nu ligt.

Prof. Dr. Frank Verhulst dank ik voor zijn rol als wegbereider van dit proefschrift, voor zijn inspirerende aanwezigheid, voor zijn talent om mij op de juiste momenten te stimuleren én mijn gang te laten gaan, voor de ruimte die hij bood en biedt om mij als onderzoeker te ontplooien, voor zijn loyale steun en voor zijn kameraadschap. Zijn onverzettelijke wil en nooit aflatende ijver om de afdeling kinder- en jeugdpsychiatrie van het Sophia Kinderziekenhuis tot een werkelijk academische afdeling te smeden zijn voor mij een bron van inspiratie geworden.

I want to thank Prof. Dr. Tom Achenbach for helping me choosing the very worthwile subject of this dissertation, and for his in many ways exemplary work on the assessment and taxonomy of child and adolescent psychopathology. I feel honored to offer this small contribution to his huge scientific endeavor.

De leden van de kleine promotiecommissie, Prof. Dr. H. van Engeland, Prof. Dr. R. Giel, en Prof. Dr. J. Passchier zeg ik dank voor hun bereidheid om een deel van hun kostbare tijd vrij te maken ter lezing en beoordeling van het manuscript.

Het onderzoek waarover in dit proefschrift gerapporteerd werd was het werk van velen. Ik wil hen danken voor hun grote inzet, hun enthousiasme, hun bereidwillige medewerking en de plezierige samenwerking, die ik in verschillende fasen van het project mocht ervaren.

In de eerste plaats denk ik daarbij aan de ouders van de kinderen, die de tijd namen om een van de interviewsters thuis te ontvangen en vragenlijsten in te vullen. Hen wil ik bedanken voor alle vertrouwelijke informatie, die de basis vormt voor een belangrijk deel van dit proefschrift.

Tati Boeken Kruger-Mangunkusumo, destijds kinderarts bij het Kruiswerk Zuid-Holland, was in het begin van het onderzoek behulpzaam bij de werving van financiële middelen en verschafte de toegang tot het adressenbestand van het Entregister Zuid-Holland. Het Entregister Zuid-Holland en de GGD Rotterdam zijn wij erkentelijk voor de bereidwilligheid en het vertrouwen waarmee zij de namen en adressen van de kinderen in het

onderzoek beschikbaar stelden.

Sylvia Rasenberg, Marije van Grol, Lianne Schouwenberg, en Désirée van de Meer zijn er in geslaagd in relatief korte tijd 420 ouders van kinderen in de algemene bevolking te interviewen en de gegevens rijp te maken voor analyse. Alice Molderink, Marike Jellema, Carolien Strick van Linschoten, en Trudy Diteweg deden hen dat na bij de gegevensverzameling tijdens de follow-up twee jaar later. Voor een groot deel per openbaar vervoer bezochten deze acht psychologen en pedagogen 105 verschillende gemeenten in Zuid-Holland, van Achthuizen tot Zwammerdam. Aan hun volharding is het te danken dat de respons boven de 90 procent kwam en bleef.

Emilie Kuipérie-van Houten, Lianne Schouwenberg, Jacqueline van Gent en Rhea Wilken verzamelden een deel van de vragenlijst- en dossiergegevens over de groep kinderen verwezen naar de instellingen voor de Geestelijke Gezondheidszorg.

In deze instellingen gaven vele personen genereus hun steun bij de benadering van de ouders en verzorgers van de aangemelde kinderen en in goed vertrouwen de toegang tot de dossiers. Ik wil met name bedanken L. van de Ban (Pedagogisch Consultatiebureau Rotterdam), H. van den Berg (Riagg RNO), G. Couturier (MKD Teesinkweide), Dr. G. Berden (Kinderpsychiatrie SKZ), Dr. D. Deboutte (Riagg Rotterdam Zuid), T. Doreleijers (Kinderpsychiatrie JKZ), H. Gerritzen (Kinderpsychiatrie AZU), E. Richter (Riagg RNW), T. van Ruhe (MKD De Kleine Plantage), Prof. Dr. J. Sanders-Woudstra (Kinderpsychiatrie AMC), en Dr. N. Weisglas-Kuperus (MKD Hoogvliet). Zonder hun bereidwillige medewerking hadden belangrijke delen van het onderzoek niet kunnen worden uitgevoerd.

Jan van der Ende heeft het reken- en schrijfwerk aanzienlijk verlicht door mij regelmatig van nieuw en steeds geavanceerder rekentuig te voorzien en door te informeren of dat ook beviel.

Dr. Edwin van den Oord was mijn metgezel op de lastigste helling in het parcours, namelijk de analyses voor paragraaf 4.1. Hij heeft gelukkig heel veel kopwerk verricht.

Gerri-Janne de Kwant nam mij op de haar eigen zorgvuldige wijze een belangrijk deel van het werk aan de bijlagen bij het proefschrift uit handen.

Judith Herzberg wil ik danken voor haar toestemming om in dit proefschrift het gedicht op te nemen dat zo treffend de wereld van het jonge kind weerspiegelt én, maar zo zal ze het nooit bedoeld hebben, de positie van de kinderpsychiatrisch onderzoeker.

Curriculum Vitae

Hans Koot (geboren op 29 december 1950 te Valkenswaard) behaalde in 1970 het diploma gymnasium α aan het Hertog Jan College te Valkenswaard. Hierna volgde hij de opleiding Verpleegkunde B in het psychiatrisch centrum St. Willibrord te Heiloo, waar hij tot 1973 als psychiatrisch verpleegkundige in dienst was. Van 1973 tot 1976 werkte hij als sociotherapeut bij de psychotherapeutische gemeenschap De Oosthoek te Limmen en bij de Prof. Pompekliniek voor forensische psychiatrie te Nijmegen. Van 1977 tot 1983, tijdens zijn studie psychologie, was hij als onderzoeksassistent werkzaam aan de Katholieke Universiteit Nijmegen bij de vakgroepen Ontwikkelingspsychologie, Klinische psychologie, en Psychologie van de Massacommunicatie, bij het Nijmeegs Universitair Huisartsen Instituut, en bij het Instituut voor Toegepaste Sociologie. In 1983 legde hij aan deze universiteit -cum laude- het doctoraalexamen Psychologie af, met als hoofdrichting Ontwikkelingspsychologie.

Van 1983 tot 1988 was hij als junior wetenschappelijk medewerker verbonden aan de vakgroep Ontwikkelingspsychologie van de KUN. De belangrijkste taken waren daar de uitvoering van onderzoek naar de vroegkinderlijke sociaal-emotionele ontwikkeling; uitvoering van pre- en postkandidaats onderwijs aan studenten psychologie; en deelname in het dagelijks bestuur van de vakgroep. In dezelfde periode was hij docent aan de opleiding Pedagogiek van de Katholieke Leergangen te Sittard en in de PAOS-cursus 'Psychologen en pedagogen in de zwakzinnigenzorg'.

Van 1988 tot 1993 was hij aangesteld als wetenschappelijk onderzoeker bij de afdeling Kinder- en Jeugdpsychiatrie van het Sophia Kinderziekenhuis / Erasmus Universiteit te Rotterdam op een subsidie van het Stimuleringsprogramma Gezondheidsonderzoek (SGO). Voornaamste taken tijdens deze aanstelling waren de uitvoering van onderzoek naar de vaststelling, het vóórkomen en het verloop van probleemgedrag bij kinderen en jongeren; consultatie en advisering binnen en buiten het ziekenhuis; en onderwijs aan studenten geneeskunde en psychiaters in opleiding. In deze periode deed en begeleidde hij onderzoek naar de validiteit en bruikbaarheid van de Child Behavior Checklist voor 2-3-jarigen, waarvan dit proefschrift verslag doet; het verloop van probleemgedrag bij kinderen en jeugdigen in de ambulante GGZ en in de algemene bevolking; probleemgedrag bij kinderen en jeugdigen in de semi-residentiële jeugdhulpverlening; het functioneren van ex-cliënten uit de therapeutische gezinsverpleging; de relatie tussen gezinskenmerken en probleemgedrag (met subsidie van NWO); de ontwikkeling van een instrument voor gezinsproblemen; en het psychosociale functioneren van somatisch zieke kinderen.