THE STRENGTHS AND DIFFICULTIES QUESTIONNAIRE

Among Finnish School-Aged Children and Adolescents

by

Merja Koskelainen

From the Department of Child Psychiatry
University of Turku
Turku Finland

Supervised by:
Docent Andre Sourander, MD, PhD
Department of Child Psychiatry
University of Turku
Turku Finland
and
Professor Pekka Niemi
Department of Psychology
University of Turku
Turku Finland

Reviewed by:
Docent Hanna Ebeling
Department of Child Psychiatry
University of Oulu
and
Professor Timo Ahonen
Department of Psychology
University of Jyväskylä

Opponent:
Research professor Matti Rimpelä
National Research and Development
Center for Welfare and Health
(STAKES)

ABSTRACT

Merja Koskelainen

The Strengths and Difficulties Questionnaire (SDQ-Fin) among Finnish children and adolescents

Department of Child Psychiatry, University of Turku, Turku, Finland

Objective: The psychometric properties of The Strengths and Difficulties Questionnaire (SDQ-Fin), a Finnish version of a brief screening instrument were studied. Emotional and behavioural problems of 7- to 15-year-olds measured by the SDQ were reported, as well as the occurrence of self-reported eating disturbance symptoms and alcohol use among adolescents.

Methods and samples: The cross-sectional school survey included 25 items of the SDQ-Fin, items about eating disturbance, alchol use and child psychiatric help-seeking. The study consists of three community samples: 1. The SDQ-Fin parent (n = 703) and teacher (n = 376) versions of 7 – 12 –year-olds, and self-report versions (n = 528) of 11 – 16 years-olds were obtained, and 2. the parent (n = 81) and self-report versions of 15-16 –year olds (n = 129) were obtained in Laitila and Pyhäranta. 3. The self-report versions of 13 – 16 – year-olds (n = 1458) in Salo and Rovaniemi were obtained.

Results: The psychometric properties of the SDQ-Fin were for the most part comparable with the other European SDQ research results. The internal consistency (Cronbach's alpha = 0.71 in all informants' reports) and inter-rater reliability (between the pairs of reports r = 0.38 - 0.44) were adequate. The concurrent validity (r = 0.75 between the SDQ and the CBCL total scores; r = 0.71 between the SDQ and the YSR total scores) was sufficient. Factor analysis of the SDQ self-report generally confirmed the postulated structure for girls and boys, except for the conduct problems scale of boys, which was fused with emotional symptoms and with hyperactivity. The response rates, means and cut-off points of the SDQ self-report scores were similar to those found, e.g. in Norway and in Britain. A high level of psychological problems, especially emotional and conduct problems and hyperactivity-inattention, were associated with high level of eating disturbance symptoms and alcohol use.

Conclusion: The results showed that the psychometric properties of the SDQ-Fin are adequate and provide additional confirmation of the usefulness of the SDQ-Fin for, e.g. screening, epidemiological research and clinical purposes.

Keywords: The Strengths and Difficulties Questionnaire, the SDQ, screening, child and adolescent psychiatry, eating disturbance, adolescent self-reported alcohol use

TIIVISTELMÄ

Merja Koskelainen

The Strengths and Difficulties Questionnaire (SDQ-Fin) among Finnish children and adolescents

Lastenpsykiatria, Turun yliopisto, Turku

Tavoitteet: Vahvuudet ja vaikeudet –kyselylomake (the Strengths and Difficulties Questionnaire, SDQ-Fin) on tätä tutkimusta varten laadittu suomenkielinen käännös lyhyestä, laajalti käytetystä kyselylomakkeesta. Tutkimuksen tavoitteena oli kartoittaa SDQ-Fin -kyselylomakkeen psykometrisia ominaisuuksia, reliabiliteettia ja validiteettia suomalaisella aineistolla. Tavoitteena oli myös kuvata lomakkeen avulla kartoitettuja tunne-elämän ja käyttäytymisen ongelmia sekä tutkia näiden yhteyksiä nuorten psykososiaalisiin ongelmiin: syömisongelmiin ja alkoholinkäyttöön**.**

Menetelmät ja otokset: Koululaiskysely sisälsi SDQ-Fin lomakkeen 25 kysymystä sekä kysymyksiä lastenpsykiatriseen hoitoon hakeutumisesta, syömisongelmista ja alkoholinkäytöstä. Tutkimus tehtiin kolmesta otoksesta: Laitilassa ja Pyhärannassa ensinnäkin 7-12 –vuotiaiden vanhemmat (n = 703) ja opettajat (n = 376) sekä 11-16 –vuotiaat nuoret itse (n = 528) täyttivät lomakkeet ja toiseksi 15-16 –vuotiaat nuoret (n = 129) ja heidän vanhempansa (n = 81) täyttivät kyselylomakkeet. Salossa ja Rovaniemellä 13-16 –vuotiaat nuoret (n = 1458) täyttivät kyselylomakkeen.

Tulokset: SDQ-Fin kyselylomakkeen psykometriset ominaisuudet olivat suurelta osin vertailukelpoisia kansainvälisten tutkimusten tulosten kanssa. SDQ-Fin kokonaisongelmien asteikon sisäinen yhtenäisyys (Cronbachin alpha = 0.71), arvioitsijoiden välinen reliabiliteetti (arvioitsijaparien välillä r = 0.38 – 0.44) sekä samanaikaisvaliditeetti (kokonaisongelmien pistemäärissä SDQ-Fin ja CBCL välinen korrelaatio = 0.75; SDQ-Fin ja YSR välinen korrelaatio = 0.71) olivat riittävän hyviä. Faktorianalyysin tulokset vahvistivat tyttöjen ja poikien itse täyttämien lomakkeiden osalta oletetun rakenteen, lukuun ottamatta poikien osalta käytöshäiriön faktoria, jonka kysymykset korreloivat vahvemmin tunne-elämän ja tarkkaavaisuusongelmien faktoreiden kanssa. Nuorten itsensä täyttämissä lomakkeissa ongelmien prevalenssi, keskiarvot ja katkaisupisteet olivat samansuuntaisia kuin esim. norjalaisissa ja brittiläisissä tutkimuksissa. Lisäksi nuorten kuvaamien ongelmien määrä, erityisesti tunne-elämän, käytöshäiriön ja tarkkaavaisuusongelmien määrä, korreloi syömisongelmien ja alkoholin käytön määrän kanssa.

Johtopäätökset: Tutkimuksen tulokset osoittivat, että Vahvuudet ja vaikeudet –kyselylomakkeen psykometriset ominaisuudet ovat suurelta osin hyvät ja se on käyttökelpoinen menetelmä seulontaan, epidemiologiseen tutkimukseen ja kliiniseen käyttöön.

Avainsanat: Vahvuudet ja vaikeudet – kyselylomake, SDQ, seulonta, lasten ja nuorisopsykiatria, syömisongelmat, nuorten alkoholikäyttö

CONTENTS

ABBREVIATIONS

SDQ	Strengths and Difficulties Questionnaire
SDQ-Fin	Strengths and Difficulties Questionnaire, Finnish version
Rutter A1	Rutter's Parent Questionnaire A1
Rutter B2	Rutter's Teacher Questionnaire B2
CBCL	Child Behaviour Checklist
YSR	Youth Self Report
TRF	Teacher Report Form
CDI	Children's Depression Inventory
RCMAS	Revised Children's Manifest Anxiety Scale
EDI	Eating Disorder Inventory
DAWBA	Development and Wellbeing Assessment
PACS	Parental Account of Child Symptoms
CABA	Child and Adolescent Burden Assessment
PBQ	Prosocial Behaviour Questionnaire
ROC	Receiver Operating Characteristics
CORC CAMHS	Outcome Research Consortium
CAMHS	Collaboration between Child and Adolescent Mental Health Services
ESCAP	European Society for Child and Adolescent Psychiatry

LIST OF ORIGINAL PUBLICATIONS

I   Koskelainen M, Sourander A, Kaljonen A (2000) The Strengths and Difficulties Questionnaire among Finnish school-aged children and adolescents. European Child and Adolescent Psychiatry 9(4): 277 – 284.

II   Koskelainen M, Sourander A, Vauras M (2001) Self-reported strengths and difficulties in a community sample of Finnish adolescents. European Child and Adolescent Psychiatry 10(3): 180 – 185.

III   Koskelainen M, Sourander A, Helenius H (2000) Dieting and weight concerns among Finnish adolescents. Nordic Journal of Psychiatry 55: 427 – 431.

IV   Koskelainen M, Sourander A (2001) Self-reported alcohol use and behavioural problems among Finnish adolescents. Psychiatria Fennica 32: 57 – 66.

V   Obel C, Heirvang E, Rodriquez A, Heyerdahl S, Smedje H, Sourander A, Gudmundsson O, Clench-Ass J, Christesen E, Heian F, Mathiesen KS, Magnusson P, Njardvik U, Koskelainen M, Rønning JA, Stormark KM, Olsen J. (2004) The strengths and difficulties questionnaire in the Nordic countries. European Child and Adolescent Psychiatry (Suppl 2) 13:II/32 – II/39.

1.   INTRODUCTION

1.1   Screening questionnaires in child psychiatry

The Strengths and Difficulties Questionnaire is a new, brief behavioural screening questionnaire for 4 – 16-year-olds. It was originally developed by Robert Goodman (since 1997) from the Department of Child and Adolescent Psychiatry, King's College, London. It consists of 25 items in five different domains: conduct problems, emotional symptoms, hyperactivity-inattention, peer problems and prosocial behaviour, and it contains both positive and negative behavioural traits. There are similar versions for different informants: parents, teachers and 11-16-year-old children and adolescents themselves. There is also another version for the parents and the day care teachers of 3- to 4-year old children. The SDQ can be used by various mental health professionals and for different purposes, e.g. for screening, for epidemiological research, for clinical assessment and for evaluating intervention outcome. Since the questionnaire can be completed in about five minutes, a positive effect on its acceptance by the responding informants can be expected, thus leading to low rates of refusal and missing answers.

In the child psychiatric epidemiology research, questionnaires are usually used in the first stage of surveys, and are followed by clinical interviews in the second stage. The better the first screening stage can be conducted, the easier and more accurate it becomes to apply the results and to choose the criteria for the clinical sample (Kresanov et al. 1998; Fombonne 1991). Screening in the first phase of a two-stage community survey should be achieved with sufficient efficiency. However, second-stage interviews are more difficult and more costly to carry out. Thus, sampling the screen-positive group should result in an elevated probability of selecting a case for further investigation, in order to make optimal use of interviewing resources. (Fombonne 1991). Thus, there is a need for further development of efficient and accurate assessing methods in child psychiatric research based on current theories of children's emotional and behavioural problems.

Besides the SDQ, there are only few questionnaires to assess a broad range of children's mental health problems. The Rutter questionnaires and the Achenbach questionnaires, the Child Behaviour Check List (CBCL), the Teacher Report Form (TRF) and the Youth Self Report (YSR) are well known and widely used. All have their advantages as well as their disadvantages. The Rutter questionnaires (McGee et al 1985) for screening children's emotional and behavioural problems are long-established and highly respected behavioural screening instruments that have proved valid and reliable in many contexts (Elander & Rutter 1996). The Rutter parent version (A2-form) was developed to discriminate between potentially healthy and potentially disturbed children in a normal population. It includes 31 items, scored 0 - 2. Items concern the child's health, abilities

and habits (e.g. speech development, eating and sleeping), and the child's behaviour. It can be used to generate scores for total deviance (ranging from 0 to 62), conduct problems, emotional symptoms and hyperactivity, as well as predictions of the type of psychiatric disorder (Rutter 1967; Goodman 1994; Kresanov et al. 1998). The Rutter B2-form for teachers asks about the child's behaviour at school. It consists of 26 items, concerning the child's health, habits and behaviour. Of these items, 23 are the same as on the parent's form, while the others are about situations at school. The Rutter questionnaires have been used in many epidemiological studies (e.g. Rutter 1967 and 1970; Almquist et al. 1991; Puura et al.1995).

Since the 1960s, the Rutter questionnaires have been widely used but they show their age in some ways. Many areas of contemporary interest, including concentration, impulsivity, reflectiveness, having friends, being victimized and acting prosocially, are poorly covered. Furthermore, all items are about negative traits, whereas the recent trend has been to emphasize children's strengths, not just their deficits (Goodman 1997).

The Child Behaviour Checklist (CBCL) is a standardized measure of behavioural problems for parents of 6 – 18-year-olds (and 1.5 – 5 –year-olds), the Teachers' Report Form (TRF) is for teachers (a report form for day-care workers), and the Youth Self Report (YSR) for the 11 – 18- year-old children and adolescents themselves (Achenbach 1991a; 1991b; 1991c). CBCL and TRF consist of 118 behaviour items, each scored from 0 to 2. They give a total behaviour problem score and two broad sub-scores: Externalizing, which includes aggression, conduct disorder, delinquent behaviour, hyperactivity and cruelty, and Internalizing, which includes depression, anxiety, withdrawal and somatizing. The CBCL also includes the Social Competence scale, while the TRF includes academic and adaptive functioning.

The Achenbach questionnaires are among the most common epidemiological instruments used today in research and in clinical assessment of children's and adolescents' emotional and behavioural problems. The CBCL, YSR and TRF are designed to collect data on a wide variety of behaviours that are of clinical concern. For screening and research purposes, the Achenbach questionnaires seem less useful as they are quite long and contain many items that are not relevant to the majority of children. Although the CBCL is substantially longer and therefore more time-consuming to complete than the Rutter questionnaires, it also has its advantages, e.g. there is a version for self-report. The validity and reliability of the CBCL and the YSR are well documented internationally, e.g. in the USA (Achenbach and Edelbrock 1983; Achenbach 1991a; 1991c) and in Europe, e.g. in the Netherlands (Verhulst et al 1985), and in Finland (Almquist et al 1988).

Contrary to the few questionnaires of broad range, there are many questionnaires concerning specific domains of mental problems, e.g. ADHD (the Conners Rating Scales-Revised, Conners 2001; Hudziak et al. 2005), depression (the Children's Depression

Inventory (CDI), Kovacs 1981), anxiety (the Revised Children's Manifest Anxiety Scale (RCMAS), Reynolds and Richmond 1978), and eating disorders (the Eating Disorder Inventory (EDI), Garner DM 1991; Garner et al. 1982; Keski-Rahkonen et al. 2006).

2. THE STRENGTHS AND DIFFICULTIES QUESTIONNAIRE

2.1 Development of the SDQ by Robert Goodman

The Strengths and Difficulties Questionnaire (SDQ) is an instrument for examining emotional and behavioural problems of children and adolescents (Goodman 1997; 1999; Goodman et al. 1998; Goodman and Scott 1999). The reliability and validity of the SDQ have been investigated since 1997 in the United Kingdom by Goodman and colleagues. Before that, Goodman (1994) had modified the Rutter parent questionnaire by including extra items on children's strengths. When using the Rutter parent and teacher questionnaires in an assessment battery in an epidemiological study, he noticed in a pilot study that many parents and teachers found the focus of Rutter items disconcerting (Goodman 1994). All items were about undesirable traits, and respondents often felt that they wanted to identify the child's strengths as well as weaknesses.

In order to make the questionnaire more acceptable for respondents, and thereby increase the response rate, Goodman (1994) added extra items on children's strengths. The original order of the Rutter items was maintained without any change in wording and the additional 19 items were placed in pseudo-random manner in the main section of the questionnaire. Ten of the 19 added items concerned prosocial behaviour. Four of these items were taken from and three items were similar to the Prosocial Behaviour Questionnaire (Weir and Duveen 1981), but the final three prosocial items were novel. Other additional items were about positive traits, which can be thought of as the opposite of psychopathological traits (e.g. additional items on the hyperactivity scale in positive wording: "sees tasks through to the end"; "thinks things out before acting", "concentrates well") and neutral or undesirable traits that were not well covered by the Rutter parent questionnaire (e.g. "acts rushly and impulsively", "gets on better with adults than with other children"). The structure of the developed questionnaire was studied by principal component analysis. The rotated solution with six factors was easiest to interpret and made the greatest clinical sense: hyperactivity-inattention, prosocial behaviour, conduct problems/oppositionality, somatic/developmental, internalization, peer relationships. The internalization factor was named emotional symptoms, and the somatic/developmental (soils, wets, difficulties in sleep, eating and speech) factor was excluded from the new SDQ questionnaire.

The prosocial scale was designed to cover, at the time, new ideas and interests in child psychiatry on positive and protective factors of personality. Prosocial behaviour includes interpersonal interaction and concern for others (e.g. helping, sharing, giving, co-operating, responding to distress) (Weir and Duveen 1981; Hay 1994). Goodman (1994)

found that the prosocial scale was a distinctive dimension of behaviour, not simply the opposite of antisocial, neurotic or hyperactive behaviour. This was evidenced by the factor structure and the modest or nonexistent correlations between the prosocial behaviour score and the behaviour problems scores.

The first article describing the SDQ was published by Goodman in 1997. A new screening questionnaire, the SDQ was administered along with Rutter questionnaires to parents and teachers of 403 children aged 4-16 years, drawn from dental and psychiatric clinics. Scores obtained by the SDQ and the Rutter were highly correlated; parent-teacher correlations for the two sets of measures were comparable or favoured the SDQ. The two sets of questionnaires did not differ in their ability to discriminate between psychiatric and dental clinic attenders.

The SDQ self-report version was studied by Goodman in 1998. The SDQ self-report was administered to two samples of 11-16-year-olds drawn from the community (n = 83) and a mental health clinic (n=116). The questionnaire discriminated satisfactorily between the two samples. The clinical cases were over six times more likely to have a score at or above the abnormal range (90th percentile). The inter-rater correlations between the SDQ self-, parent and teacher reports compared favourably with average cross-informant correlations in previous studies on a range of measures (e.g. Achenbach et al. 1987).

Concurrent validity of the SDQ was studied by comparing the parent-rated SDQ and the CBCL on 132 children aged 4 through 7, and drawn from psychiatric and dental clinics (Goodman and Scott 1999). The results showed that scores derived from the SDQ and the CBCL were highly correlated and equally able to discriminate psychiatric from dental cases. As judged against a semi-structured interview (The Parental Account of Child Symptoms, PACS, Taylor et al. 1986), the SDQ was significantly better than the CBCL at detecting inattention and hyperactivity, and at least as good at detecting internalizing and externalizing problems. Mothers of low-risk children were twice as likely to prefer the SDQ.

The next improvement was to extend the SDQ to be better able to detect the probable child psychiatric cases (Goodman 1999). The extended SDQ includes an impact supplement that asks wheather the respondent thinks the young person has a problem, and if so, enquires further about chronicity, distress, social impairment and burden to others. Similar versions are completed by parents, teachers and children aged 11 to 16. The validation study involved two groups of 5-15-year-olds: a community sample (n = 467) and a psychiatric clinic sample (n = 232). The two groups had markedly different distributions on the measures of perceived difficulties, impact (distress plus social impairment) and burden. Impact scores were better than symptom scores at discriminating between the community and clinical samples; however, discrimination based on the single "Is there a problem?" item was almost as good. The SDQ burden rating correlated well

(r=0.74) with a standardized interview rating of burden (Child and Adolescent Burden Assessment, CABA; Messer et al.1996). For clinicians and researchers with an interest in psychiatric caseness and the determinants of service use, the impact supplement of the extended SDQ appears to provide useful additional information without taking up much of the respondents' time.

Goodman et al. (2000) wanted to further increase the clinical usefulness and the ability of the SDQ to detect possible psychiatric cases. They developed a computerized algorithm to predict a child psychiatric diagnosis on the basis of the symptom and impact scores derived form the SDQ completed by three informants. The algorithm predicted that a disorder was probably present on the basis that the relevant symptom score was above the 95th percentile and the impact score was two or more (representing a minimum of "quite a lot" of impact in two domains or "a great deal" of impact in one domain).

The algorithm was applied to patients attending child mental health clinics in Britain (n = 101) and Bangladesh (n = 89). In both London and Dhaka, children were assigned an ICD-10 clinical diagnosis on the basis of detailed information on symptoms and resultant impairments gathered from multiple informants. These diagnoses were grouped into three broad categories to provide large enough cell sizes for analysis: hyperkinesis, conduct disorder (including oppositional disorder) and emotional disorder (including anxiety, depressive, and obsessive compulsive disorders). As a criterion, each disorder was rated as absent, borderline or present by a child psychiatrist who was blind to the SDQ results. The results show that for each disorder in each clinic, the level of agreement between clinical diagnosis and the SDQ prediction was substantial and highly significant (Goodman et al.2000).

Lately Goodman and colleages have developed a semi-structured interview, the Development and Well-Being Assessment (DAWBA; Goodman et al. 2000; Goodman 2001), and together with the SDQ, they form an integrated package of questionnaires, interviews and rating techniques designed to generate psychiatric diagnoses in 5-16-year-olds. Currently the DAWBA has been translated into Finnish by Andre Sourander and co-workers.

In the present study, the Finnish translation of the SDQ (SDQ-Fin) was used for the first time for research purposes. When the Finnish SDQ studies were conducted, the SDQ had been translated into only few languages (e.g. Swedish and German) in addition to English. Currently a growing body of research exists concerning the SDQ in different countries. Research pertinent to SDQ is conducted virtually all over the world, e.g. in Europe, in the USA, in Africa, in Australia and in Asia. It has been translated into more than 60 languages.

3.   INTERNATIONAL STUDIES WITH THE SDQ

3.1   Studies of the SDQ in Nordic countries

The SDQ was translated into the different Nordic languages between 1996 and 2003. During the past few years, the SDQ has been completed by nearly 100 000 children and adolescents in community-based studies, as well as in clinical samples (Obel et al.2004). In a supplement issue of the European Child and Adolescent Psychiatry, Obel et al. (2004) describe, in addition to more widely published research results, many still ongoing research projects and studies of which reports have been published only in the respective language, e.g. in Denmark and Iceland. In this section, examples of studies conducted in different countries are described.

In Sweden, Smedje et al. (1999; 2001) used the SDQ parent ratings from samples of 6-8-year-old and 10-year-old children (n = 900) from the general population. They studied sleep problems in the 6-8 year-olds, and according to the parent SDQ ratings, found more problematic daytime behaviour in children with sleep problems. When studying the psychometric properties of the parent-report SDQ, they found that they were adequate and that the results confirmed the postulated factor structure. To study the internal reliability of the SDQ, an informal test-retest procedure over a two-week period (n = 15) was conducted. The cut-off point according to the 90th percentile of the SDQ total scores was found to be somewhat lower than in the British study (Goodman 1997). Moreover, some significant gender differences were found; boys scored higher on the SDQ total scale and the hyperactivity scale, while girls scored higher on the prosocial scale. In a later Swedish study (Malmberg et al. 2003), a community sample (n = 263) and a child psychiatric sample (n = 230) were compared, and it was found that the extended parental SDQ discriminated well between clinical and non-clinical cases.

In Norway, several large-scale studies have been performed since the Norwegian translation of the SDQ became available in 2001. A recent report (Rønning et al. 2004) dealt with the psychometric properties of the SDQ self-report version. It is based on a study conducted in 66 primary and secondary schools (n = 4167). In their study, Rønning et al. (2004) presented the SDQ total and subscale means split by gender and age across grades from the 6th to the 10th. The SDQ scale analyses included internal consistency, and intra- and cross-scale correlations and confirmatory factor analysis. The results showed somewhat variable and questionable psychometric properties, especially concerning the factor structure, internal consistency and cross-scale correlations. Cut-off points were similar to those reported in the Finnish study (Koskelainen et al. 2000).

A large-scale Norwegian study was conducted in Akershus County, surrounding Oslo. Van Roy et al. (2006) reported the results based on the extended Norwegian SDQ self-

report version (items on impact and burden included) for children and adolescents (n = 29631). They present the SDQ self-reported total and subscale means split by gender and age, in three age groups, i.e. children and adolescents attending primary, secondary, and upper secondary school (10- to 20-year-olds). According to their results, girls reported most emotional problems and boys reported most conduct and peer problems. Combining symptoms scores and impact scores, 3.7 % of the sample was defined as a high risk group, while 8.8 % needed special attention from mental health services. Compared to Finland, North-Norway and the UK, Norwegian adolescents, both boys and girls, reported more hyperactive behaviour (Koskelainen et al.2000; Rønning et al. 2004; www.sdqinfo.com).

3.2   Studies of the SDQ in other European countries

In other European countries, numerous SDQ studies has been conducted, e.g. in Germany (Klasen et al. 2000; Woerner et al. 2004; Becker et al. 2004a and 2004b), Holland (Muris et al. 2003; Wiedenfelt et al. 2003) and the UK (Goodman 1997; Goodman 1999; Goodman 2001; Goodman et al. 1998; Goodman and Scott 1999; Goodman et al. 2000). Marzocchi et al. (2004) described the SDQ research conducted in Southern European countries. For the most part, according to the author's knowledge, these research projects are ongoing or reported only in Italian, French, Spanish, Portuguese, and they are not included here.

In Germany, Woerner et al. (2004) investigated the psychometric properties of the parent-rated SDQ; the scale structure, distribution of scale scores, cut-off scores, in addition to effects of gender, age and socio-economic status obtained from a community sample aged from 6 to 16 years (n = 930). A principal component analysis resulted in a replication of the original SDQ subscales. The internal consistency of the SDQ total difficulties scale proved to be good. All the SDQ subscales also showed satisfactory homogeneity, although coefficients were lower, particularly for the peer problems scale, than for other subscales. Only one subscale was significantly associated with age, namely hyperactivity-inattention, with older children having fewer problems than younger ones. In addition, children and adolescents with less favourable conditions scored significantly higher on the hyperactivity-inattention subscale and, somewhat less markedly, on the peer problems.

For concurrent validity purposes, Klasen et al. (2000) compared the parent-rated SDQ and the CBCL, both of which were completed by parents of 273 children drawn from psychiatric clinics (n = 163, aged 4-16) and from a community sample (n = 110, aged 12-13). The children from the community sample also filled in the SDQ self-report and the YSR. The results showed that the two questionnaires were highly correlated and equally able to distinguish between community and clinical samples. According to the parent ratings, both measures showed good discriminant validity. According to the total

score, the parent-rated SDQ was better able to distinguish between community and clinical samples than the CBCL. Furthermore, the parent-self agreement of the SDQ was comparable to the parent-self agreement measured by CBCL/YSR. In the ability to differentiate between disorders within a clinical sample, there were few differences in the performance of the two questionnaires, except in the case of hyperactivity-inattention, where the SDQ performed significantly better. However, it is noteworthy that both questionnaires were better able to differentiate between clinical and community samples than to differentiate between different types of disorders.

Becker et al. (2004a) have studied the validity and reliability of the German SDQ parent-, teacher and self-rated versions in a clinical sample. They examined children and adolescents aged from 5 to 17 years (n = 543) and correlated the results of the parent-rated and teacher-rated SDQ, as well as the CBCL/TRF with clinical diagnoses. They used confirmatory and exploratory factor analyses and found an exact replication of the original SDQ five-scale structure. Correlations between the SDQ scales and the corresponding CBCL/TRF scales showed a high degree of congruence. When they used the total difficulties score (SDQ) or the total problems score (CBCL/TRF) to predict presence of any behavioural disorder, they were able to show that the SDQ, as well as the CBCL and TRF, were equally able to differentiate between patients with and without any clinical diagnosis. Moreover, the examined subscales could also detect more specific clinical subgroups, revealing only minor differences between the two questionnaires.

The SDQ self-report was studied by Becker et al. (2004b) when data were collected from 214 child psychiatric in- and outpatients aged 11 to 17 years. Results were compared with the parent and teacher SDQ versions, the corresponding CBCL/YSR scores and the clinical diagnoses. Internal consistency was found to be sufficient, and correlations between the equivalent parent and teacher SDQ ratings were acceptable. The self-rated SDQ differentiated well between clinically defined cases and non-cases, and also detected various subcategories of psychiatric disorders within the clinical sample. In their study, the SDQ self-report significally contributed to the prediction of the diagnoses, specifically if only parent or teacher ratings were available.

In a Dutch study (Muris et al. 2003), the parent- and self-rated SDQ versions were obtained from 9- to 15-year-old children and adolescents (n = 592). Principal component analysis of the parent-rated SDQ showed that all items loaded on the predicted five factors. Furthermore, internal consistency, test-retest stability, and parent-youth agreement of the various SDQ scales were acceptable. Also the concurrent validity was found to be good. The SDQ scores correlated in a meaningful way with various parent versions of other questionnaires used in the study (e.g. CBCL, CDI-P, RCMAS-P).

In another Dutch study, Wiedenfelt et al. (2003) collected several samples of various SDQ versions. Self-report data was collected from 11- to 16-year-olds (n = 970), the

extended version of the SDQ self-report from 11- to 16-year-olds (n = 268), the SDQ parent version from 8- to 16- year-olds (n = 300), and the SDQ teacher version from 8- to 12-year-olds (n = 208). In addition, the CBCL was completed by parents and the YSR by the 11- to 16-year-olds. Their results showed that the internal consistency of the teacher-rated SDQ was good, and that of parent and self-rated SDQ acceptable. The mean inter-informant correlations of the SDQ scales were satisfactory. Concurrent validity with other measures used in their study (CBCL, YSR, CDI, RCMAS) was good.

3.3   Studies of the SDQ in the USA and Australia

In the USA, a large-scale study has been conducted by Dickey and Blumberg (2004) when parents or guardians of a national probability sample of 9574 children and adolescents aged 4-17 years filled in the SDQ questionnaire. Structural analysis of the parent-rated SDQ was made by principal component analysis, and exploratory and confirmatory factor analyses. The results showed that the predicted five-component structure was not entirely confirmed. Instead of five factors, exploratory factor analyses revealed a stable three-factor model consisting of externalization problems, internalization problems, and a positive construal factor.

Bourdon and Goodman et al. (2005) used the American SDQ parent version in a large study (n = 10367) of 4- to 17-year-olds. Scoring bands were developed to differentiate low, medium and high levels of emotional and behavioural difficulties. Children at high risk of severe difficulties were identified by three different scoring methods: high symptom scores, parental perception of definite or severe difficulties and high symptoms plus impairment. These ratings were validated, e.g. against reported service contact or use. Their results showed that the SDQ was well accepted by American parents and had good internal consistency. Normative scoring bands were similar, though not identical, to the original British bands (www.sdqinfo.com). The results of each scoring method used had a strong association with service contact or use.

In Australia, many studies of the SDQ have been conducted and reported. Hawes and Dadds (2004) conducted the first empirical study of psychometric properties of the SDQ parent version using a large community sample (n = 1359) of children (aged 4-9 years). Internal consistency, stability and external validity as well as normative data and cut-off scores were reported. The results showed moderate to strong internal reliability across all SDQ subscales. The relationships of the various SDQ scores showed adequate internal validity, while correlations between the SDQ subscales, teacher ratings and diagnostic interviews showed good external validity.

Also in Australia, Mathai et al. (2004) conducted a study in which a computerized algorithm developed by Goodman et al. (2000) was used in a community sample in

Australia to predict child psychiatric diagnoses on the basis of the symptom and impact scores derived from the SDQ parent version (n = 130) and teacher version (n = 101) of 4- to 15-year-olds, and the self-rated version (n = 38) by 11- to 16-year-olds. All participants were consecutive new referrals to a multi-diciplinary child and adolescent mental health clinic near Melbourne. The SDQ results were compared with the clinical team diagnoses, in addition to an independent clinician who examined the case notes and was blind to the SDQ scores. The agreement between the results obtained by the SDQ and the diagnoses made by the clinical team was moderate to high. Agreement between the results obtained by the SDQ and diagnoses made by the independent clinician were somewhat lower. All were statistically significant.

3.4   Studies of the SDQ in developing countries

The SDQ self-report was used in Yemen by Almaqrami and Shuwail (2004) to compare the results obtained from a community sample of school-aged children (n = 600) and a clinical sample (n = 57) of consecutive patients attending a mental health hospital in Sana, the age range being from12 to 17 years. Their results showed that the difference between the means of the SDQ total difficulties scores between the two samples was highly significant; the SDQ self-report differentiated the two samples well (ROC analysis); agreement between the diagnosis and the SDQ subscales was statistically significant and factor analysis yielded the predicted five SDQ dimensions.

Samad et al. (2005) studied the validity of the Urdu version of the SDQ in Pakistan. The study population consisted of 212 children aged 4 to 16 years attending psychiatric (n = 72) and paediatric (n = 140) outpatient clinics. The parents of these children were interviewed using the extended version of the SDQ. The psychiatric group was diagnosed (by ICD-10 criteria) by a child psychiatrist who was blind to the SDQ scores. The results showed that the Urdu version of the SDQ was able to discriminate between the two groups. The SDQ mean total difficulties score was statistically significantly higher in the psychiatric group (mean = 21.7) than in the comparison group (mean = 14.4). A choice of optimal cut-offs for the Urdu version was discussed.

Kashala et al. (2005) studied the SDQ teacher ratings among school children in Kinshasa, Congo. An epidemiological study was conducted with 1187 children, aged from 7 to 9 years, recruited from randomly selected schools. Mental health problems were assessed with the teacher version of the SDQ. Stability of the factor structure was studied by principal component analysis. The reliability was evaluated using measures of internal consistency of the SDQ scales. The results showed that factor analysis yielded the predicted five factors. Using the 90th percentile, the cut-off scores were somewhat higher than the British ones.

Mullick and Goodman (2001) studied the potential suitability of the SDQ to be used for detecting child psychiatric problems in a developing country. The adequate SDQ versions were administered to parents and teachers of 261 Bangladeshi children aged from 4 to 16 years; 99 drawn from a psychiatric clinical and 162 drawn from the community. The self-report SDQ was completed by 11- to 16-year-olds. Children from the psychiatric clinic sample were assigned psychiatric diagnoses blind to their SDQ scores. The results showed that the SDQ scores distinguished well between community and clinical samples, and also between children with different psychiatric diagnoses in the clinical sample. Clinical diagnoses were classified into three broad categories, namely hyperkinesis, conduct disorder (including oppositional disorder) and emotional disorder (including anxiety, depressive and obsessive compulsive disorder). An algorithm calculated from the SDQ multi-informant ratings scores was used to predict whether children would have any of these disorders (Goodman et al. 2000).

4.   THE SDQ AND SCREENING OF PSYCHOSOCIAL PROBLEMS AMONG ADOLESCENTS

In addition to epidemiological research, the SDQ has been used in descriptive studies with children who are considered to be at "high risk" on other parameters, including psychosocial factors and developmental and physical conditions (Vostanis 2006). In many studies, the SDQ has been used as a screening method for behavioural and emotional problems in children with other diseases. The SDQ can also be used to provide baseline clinical information in studies that use mixed quantitative and qualitative methods (Vostanis 2006).

In addition to the psychometric properties of the Finnish SDQ, we wanted to study the associations between the self-reported SDQ-Fin results and some common psychosocial problems, that is, dieting and weight concerns (eating disturbances) and alcohol use reported by the Finnish adolescents (publications III-IV). Firstly, we were interested in whether a self-reported eating disturbance was associated with a high level of emotional and behavioural symptoms, as well as sex, body mass, and regular high alcohol use of the adolescents (Koskelainen et al. 2000 a). Secondly, we were interested in the associations between self-reported high level of alcohol use and emotional and behavioural problems among Finnish adolescents (Koskelainen et al. 2001 a).

4.1   Dieting and weight concerns

Concerns about weight and the body shape are very common in adolescence. Especially girls are preoccupied with dieting and a drive for thinness (Maloney et al. 1989; Schleimer 1984; Story et al. 1991). Though most adolescent dieting may be benign (Schleimer 1984), social pressures for slimness do appear to increase the risk of eating disorders. In addition, dieting is among the most important predictors of a new eating disorder (Patton et al. 1999). Excessive weight-reduction attempts by teenagers associate with health hazards including retardation of physical growth, menstrual irregularities, weakness, persistent irritability, constipation, poor concentration, sleep difficulties and impulses to binge eat (Mallick 1983; Lifshitz and Moses 1988). Depression, anxiety and substance dependence have been found to be the most common comorbid symptoms in clinical patients with eating disorders (Braun et al. 1994).

While only a minority of adolescents with even a high level of concerns over weight and dieting are likely to develop the clinical syndromes of anorexia nervosa or bulimia nervosa, a larger group may suffer some negative health effects from inappropriate dieting and weight-loss behaviours. The definition for an eating disturbance used in this study

is less categorical than diagnostic definitions used to identify clinical cases of eating disorders. Eating behaviours may be seen as a continuum that varies from normality to extreme deviance. *Eating disorder*s may be defined by diagnostic categories, while *eating disturbance*s may be described using behavioural dimensions usually on the basis of responses to self-report questionnaires (Fombonne 1997; Patton et al. 1997).

4.2   High alcohol use

The results of a large European survey studying the use of alcohol and other drugs among 15-16-year-olds living in 35 countries showed that, in Finland, the proportion of youths who had been drinking any alcohol during the last 12 months was broadly the same as average for other European countries (80 % compared to 83 %) (Hibell et al. 2003). In most European countries more boys than girls reported using alcohol (40 times or more in lifetime) and being drunk (been drunk 20 times or more in lifetime) but in a few countries, including Finland, gender distribution was about equal (Hibell et al. 2003; Ahlström et al. 2003). The gender difference was also not significant in a recent Swedish survey (n = 5445 15-18 year-olds), sixty percent of boys and 63 % of girls had been intoxicated with alcohol at least once. However, there was a significant age difference between 9th and 11th graders in the percentage of ever been intoxicated with alcohol, with those in 11th grade having been intoxicated more often (Bränström et al. 2008). Alcohol use among adolescents has also been found to be related to age in other studies (Persson 1994; Karvonen 1995; Lintonen et al. 2000).

In the proportion of those 12-16-year-olds who have used alcohol at least once a month in Finland, changes have been small in the years 1995 - 2005 (Lintonen 2005). In the European survey, the prevalence of being drunk was, however, substantially higher among Finnish adolescents than the average among European adolescents (64 % compared to 53%) (Hibell et al. 2003). Frequent high alcohol use and especially drunkenness-oriented episodic drinking, as well as early start to alcohol use in adolescents are risk factors for future alcohol dependence (Bränström et al. 2008; Grant et al. 2006). Bränström et al (2008) found that among 9th graders, monthly intensive consumption of alcohol was associated with a positive attitude to alcohol and to socializing with norm-breaking friends. Furthermore, truancy and cheating were also predictive factors for heavy drinking among boys in 9th grade, as were perceived availability of drugs and alcohol among girls of the same age.

Both epidemiological and clinical studies have demonstrated associations between adolescents' substance abuse and comorbid emotional and behavioural problems. Negative affectivity (chronically upset, frustrated, angry) and positive attitudes toward delinquency in early adolescence were associated with alchol use disorders in later adolescence (Shoal et al. 2007). In a clinical sample, about one third of adolescents in

treatment for alcohol use disorder had an attention deficit hyperactivity disorder (ADHD; 37 % of boys and 17 % of girls). Furthermore, conduct problems and substance use behaviours have been found to be closely associated, so that adolescents, who engage, e.g. in stealing, lying and starting fights were those who had identified alcohol problems and most severe substance use profiles (Molina et al. 2002). Depressive symptoms are also found to be associated with adolescents' alcohol problems. About one fourth (24 – 27 %) of adolescents identified as depressed also had problems of heavy drinking and furthermore, about one fourth of those who had heavy drinking problems (23 – 27 %) had also depressive symptoms (Windle and Davies 1999). Rohde et al. (1996) also found rates of psychiatric comorbidity with alcohol use striking, especially for disruptive behaviour disorders, depression, drug use disorders and tobacco use. More than 80 % of adolescents with alcohol disorder had another psychiatric disorder. These results indicate that a wide range of emotional, behavioural, and social competence problems is associated with a high level of alcohol use among teenagers.

5. AIMS OF THE PRESENT STUDY

The general aims of the study were to explore the psychometric properties, reliability and validity, of the SDQ parent, teacher and self-reports, when used among Finnish children and adolescents. The second aim was to explore the descriptive properties, the means, cut-off points and symptom response rates, of the SDQ-Fin. The third aim was to compare both psychometric and descriptive properties of the SDQ-Fin with international research results. The aim was also to obtain further evidence for the validity of the SDQ-Fin by exploring its association with two psychosocial problems, eating disturbance symptoms and high alcohol use among Finnish adolescents, as well as to present the descriptive properties of these two areas of psychosocial problems. More specifically, the aims of the study were:

To explore the reliability of the SDQ-Fin

1.   by studying the internal consistency and inter-rater agreement (between parent, teacher and self-reported scores) of the SDQ-Fin scales (publication I),

2.   by studying the SDQ-Fin self-report scale structure with item correlation analysis and factor analysis (publication II).

To explore the validity of the SDQ-Fin

1.   by studying the associations between child's problematic behaviour according to the parent and the clinical range (at or above 90th percentile) of the SDQ-Fin (publication I),

2.   by studying the associations between mental health help-seeking because of a child's problematic behaviour according to the parent and the clinical range (at or above 90th percentile) according to the SDQ-Fin (publication I),

3.   by studying the concurrent validity by comparing the scores obtained, firstly with the SDQ-Fin parent report and the CBCL, and secondly with the SDQ-Fin self-report and the YSR (publication I).

The aims of the study were also

1.   to present the descriptive properties of emotional and behavioural problems of Finnish children and adolescents measured by the SDQ-Fin, that is, the parent, teacher and self-reported means, symptom response rates and cut-off points (publications I and II),

2.	to study the associations between emotional and behavioural problems described by the SDQ-Fin and self-reported eating disturbance symptoms. It was assumed that especially emotional problems (e.g. depression and anxiety), as well as alcohol use, would have a high correlation with eating disturbance symptoms (publication III),

3.	to study the associations between emotional and behavioural problems described by the SDQ-Fin and self-reported alchol use by adolescents. It was assumed that especially externalizing symptoms, conduct problems and hyperactivity-inattention, but also emotional (depressive) symptoms, would have a high correlation with high alcohol use (publication IV),

4.	to compare the descriptive and psychometric results obtained by the SDQ-Fin with the results of international studies. It was assumed that the Finnish results are in line, especially with other European and Nordic results (publication V).

6. PARTICIPANTS AND MEASURES

6.1 Ethics

For the studies conducted at schools in Laitila, Pyhäranta, Salo and Rovaniemi, the study designs were approved by the school authorities. In the first sample, in Laitila and Pyhäranta, where the respondents were to answer in an identifiable manner in order to study the cross-informant correlations and where a part of the children were from primary schools, all parents of the participating children were sent a letter asking their permission for their children to participate in the study during a school lesson, in addition to the parental version of the SDQ to be filled in at home. Those parents who did not answer the first letter, were sent a second letter and if that letter remained unanswered, they were telephoned in order to get as many respondents as possible.

The second sample, also in Laitila and Pyhäranta, included respondents from secondary school (9th grade) who completed both the SDQ self-rated version and the YSR problem items anonymously during a school lesson. After that, the questionnaires were gathered in a sealed envelope and sent to the researchers. The parents of the pupils completed the SDQ parent-rated version and the CBCL also anonymously at home, and the questionnaires were sent by post to the researchers.

For the samples in Salo and Rovaniemi, the study design was also approved by the school authorities and the adolescents (7th and 9th graders) filled in the questionnaire in an anonymous way during a school lesson. After that, the questionnaires were gathered in a sealed envelope and sent to the researchers.

6.2 Participants

6.2.1 Laitila and Pyhäranta samples (publication I)

The study in Laitila and Pyhäranta was conducted during April 1998. It included all the children and adolescents, who were living in those two suburban/rural municipalities, (total population 11200) in south-west Finland and who were attending 1st, 3rd and 5th grades of primary school and 7th and 9th grades of secondary school. Only severely mentally retarded children were excluded from the study. The mean age of the children in the 1st grade was 7.8 years (sd 0.4), in the 3rd grade 9.9 years (sd 0.3), in the 5 th grade 11.9 years (sd 0.4), in the 7th grade 13.8 years (sd 0.3), and in the 9th grade 15.8 years (sd 0.4). Of the children, 48 % were boys and 52 % girls. Boys and girls were evenly distributed across the grades and the two municipalities.

The parental version of the SDQ was sent to the whole sample to be filled in at home (1st, 3rd, 5th, 7th and 9th graders; n = 703; boys 364/703, girls 339/703). The SDQ self-report was filled in by the 3rd, 5th, 7th and 9th graders (n = 528; boys 275/528, girls 253/528) in the classroom. The SDQ teacher report was filled in by the teachers of the children attending 1st, 3rd and 5th grades (n = 376; boys 180/376, girls 196/376). The teachers of the secondary school were asked to take part in the study but they refused because they thought they were not well enough informed about the children's behaviour to fill in the questionnaires (Table 1).

The parents of the children in primary and secondary school filled in the SDQ. Of the parent-report SDQ, 96 % were completed (703/735) and included in the statistical analysis. Of the children and adolescents, 89 % (534/601) had permission from their parents to participate in the study, and 99 % of them (528/534) completed the self-report appropriately. The teachers of one primary school with 33 pupils refused to participate in the study and nine more children did not have permission to participate, but all the other teacher reports at primary school (376/418) were completed and included in the statistical analysis.

The second study in Laitila and Pyhäranta was conducted in November 1998. It included all the pupils who were in the 9th grade. The adolescents completed both the SDQ self-rated version and the YSR problem items anonymously during a school lesson. The response rate was 94% (129/137). Eight questionnaires had to be excluded because the pupils were not present on the days the questionnaires were distributed, or they were inappropriately filled in. The SDQ informant-rated version and the CBCL problem items were sent to all parents of adolescents attending 9th grade. The parents also completed the questionnaires anonymously. Of the parents, 59 % (81/137) returned both questionnaires appropriately completed (Table 1).

6.2.2 *Salo and Rovaniemi samples (publications II – IV)*

The study in the cities of Salo (total population 23 655) situated in the southern part of Finland and Rovaniemi (total population 35 631) in North Finland included all children attending the 7th grade (age range 13 – 14 years) and the 9th grade (age range 15 – 17 years). Only children attending classes for the handicapped or mentally retarded were excluded from the study. The study was conducted during the last week of March, 1998. Students filled in the questionnaire anonymously during a school lesson and the questionnaires were collected in a sealed envelope. Altogether 1488 (90.7%) questionnaires were returned out of a possible 1641 7th and 9th graders living in Salo and Rovaniemi (Table 1). Most of the non-responders were not present on the days the questionnaires were distributed at the respective schools. Only 30 questionnaires were discarded because they were incomplete or inappropriately filled in. A total of 1458 (88.8 %) questionnaires were included in the statistical analysis. The mean age (sd) of boys

was 14.4 (1.1) and of girls 14.4 (1.1), of 7th graders 13.4 (0.5) and of 9th graders 15.4 (0.5). The respondents were evenly distributed between the sexes (49.7 % girls and 50.3 % boys) and across the grades (49.9 % 7th graders and 50.1 % 9th graders).

Table 1. Participants and study methods in Laitila and Pyhäranta, and in Salo and Rovaniemi.

Cities, samples	1. Laitila, Pyhäranta		2. Laitila, Pyhäranta	3. Salo, Rovaniemi
Publication	I		I	II, III, IV
School	Primary	Secondary	Secondary	Secondary
Grades	1., 3., 5.	7., 9.	7., 9.	7., 9.
Informants				
Parent (P),	P, n = 397	P, n = 306	P, n = 81/137	
(grades 1, 3, 5, 7, 9)				
Teacher (T),	T, n = 376			
(grades 1, 3, 5)				
Self (S), (grades 3,5,7,9)	S, n = 250	S, n = 278	S,n = 129/137	S, n = 1458/1641
Methods	SDQ parent-rated, SDQ self-rated, mental health help-seeking, global problem item, background information		SDQ parent-rated SDQ self-rated CBCL, YSR	SDQ teacher-rated, SDQ-self rated, dieting behaviour scale, alcohol use, background information

6.3   Measures

6.3.1 The SDQ-Fin parent, teacher and self-report (publications I – IV)

The school surveys included 25 items from the Strengths and Difficulties Questionnaire (SDQ-Fin), which (Goodman 1997; Goodman et al. 1998) includes 25 items scored 0 for "not true", 1 for "somewhat true" and 2 for "certainly true". Five of the items are worded positively and scored in the opposite direction (2 for "not true" etc.). Ten of the items are worded in a manner reflecting strengths, 14 items reflecting difficulties, while one item may be considered neutral but is scored as a difficulty item on the peer problems subscale. The 25 SDQ items are divided into 5 scales of 5 items: the hyperactivity-inattention scale, the emotional symptoms scale, the conduct problem scale, the peer problem scale and the prosocial behaviour scale. The scores of hyperactivity-inattention, emotional symptoms, conduct problems and peer problems can be summed to generate a total difficulty score ranging from 0 to 40. The prosocial scale gives a score for positive prosocial behaviour (Goodman 1997), and this sum is not included in the total difficulties score. In the present study, the cut-off point for clinical range (at or above 90th percentile for the SDQ total

scores) was used as recommended by Goodman (1997). The questionnaire was carefully translated into Finnish by the authors of this study and back- translated by a professional translator. The differences were discussed and item wording decided by the authors. The second, revised version translated by professor Seija Sandberg can be found at the SDQ website (www.sdqinfo.com) and should be used in future research.

The informant-rated version of the SDQ (publication I) can be completed by either parents or teachers of children and adolescents aged between 4 and 16 (there is an additional version for parents of 3-4-year-olds). This informant-rated version has recently been shown to function as well as the Rutter questionnaires (Goodman 1997). There is also evidence that the parent-rated SDQ correlates highly with the CBCL despite being much quicker to complete (Goodman and Scott 1999).

The self-report version of the SDQ (publications I – IV) is designed for self-completion by children and adolescents aged between 11 and 16. The 25 items cover the same attributes as the informant-rated SDQ. For most items, the only difference between the informant-rated and the self-rated version is a grammatical change from third person to first person. Since the informant-report version was designed for use with children as young as four, some of the items were not suitable for teenagers and the wording of the self-report version was modified accordingly (Goodman et al. 1998). The close correspondence of the two versions was designed to increase the comparability of the ratings obtained from different informants. The cut-off points of the 90th percentile of the SDQ scores were used to suggest caseness as in the British and Swedish studies (Goodman 1997; Goodman et al. 1998; Smedje et al.1999). Goodman presented provisional cut-off scores for an "abnormal" range so that they covered roughly 10 % of children in the community (Goodman 1997; Goodman et al. 1998).

6.3.2 The help-seeking variables (publication I)

The help-seeking variables were: (a) Use of child mental health services according to parent. The parents were asked: "Have you considered using child mental health services for examination or treatment because of the child's emotional or behavioural problems." The parents selected from three alternatives: "No" was rated 0, "we have considered" was rated 1, and "we have used" was rated 2. An additional item (b) about the parents' view of whether the child has significant emotional or behavioural difficulties: "Overall do you think that your child has difficulties in one or more of the following areas: emotions, concentration, behaviour or being able to get on with other people?" The parents selected the appropriate statement on a scale from 0 to 3. The answer "no difficulties" was rated 0, "minor difficulties" was rated 1, "definite difficulties" was rated 2 and "severe difficulties" was rated 3. In the statistical analysis of the present study the child was defined to have problematic behaviour if the parent rated 1 – 3 (on publication I).

6.3.3 The Child Behaviour Checklist (CBCL) and the Youth Self Report (YSR) (publication I)

The Child Behaviour Checklist (CBCL), a parent-report questionnaire (Achenbach 1991a) consists of 118 behaviour items, each scored on a 3-point scale. This gives a total behaviour problem score and two broad subscores, externalizing and internalizing. The externalizing score includes problems such as aggression and delinquent behaviour, while the internalizing scale includes anxiety, depression, withdrawal and somatizing. In addition to these two scales, the total problem score includes attention, social and thought problems. The CBCL also includes a social competence scale, which was not used in this study. The Youth Self Report (YSR) is a self-report version for adolescents aged between 11 and 18 years (Achenbach 1991c). It has the same kind of structure as CBCL, apart from the items being in the first person (publication I).

Table 2. The content of the SDQ items by subscales in English and in Finnish, and the order of presentation in the questionnaire sheet.

Prosocial scale	
1. I try to be nice to other people. I care about their feelings.	1. Pyrin olemaan ystävällinen muita ihmisiä kohtaan. Otan muiden tunteet huomioon.
4. I usually share with others (food, games, pens).	4. Jaan mielelläni tavaroitani (ruokaa, pelejä, kyniä jne.) toisten kanssa.
9. I am helpful if someone is hurt, upset or feeling ill.	9. Tarjoudun auttamaan, jos joku loukkaa itsensä, on pahoilla mielin tai huonovointinen.
17. I am kind to younger children.	17. Olen kiltti pienempiäni kohtaan.
20. I often volunteer to help others (parents, teachers, children).	20. Tarjoudun usein auttamaan muita (vanhempiani, opettajia, muita lapsia).
Hyperactivity scale	
2. I am restless, I cannot stay still for long.	2. Olen levoton, en pysty olemaan kauan hiljaa paikoillani.
10. I am constantly fidgeting or squirming.	10. Olen jatkuvasti hypistelemässä jotain tai kiemurtelen paikoillani.
15. I am easily distracted. I find it difficult to concentrate.	15. Häiriinnyn helposti. Minun on vaikea keskittyä.
*21. I think before I do things.**	*21. Harkitsen tilanteen ennen kuin toimin.*
25. I finish the work I'm doing. My attention is good.	*25. Saatan tehtävät loppuun. Olen hyvin pikäjänteinen.*
Emotional symptoms scale	
3. I get a lot of headaches, stomach-aches or sickness.	3. Kärsin usein päänsärystä, vatsakivusta tai pahoinvoinnista.
8. I worry a lot.	8. Murehdin monia asioita.
13. I am often unhappy, down-hearted or tearful.	13. Olen usein onneton, mieli maassa tai itkuinen.
16. I am nervous in new situations. I easily lose confidence.	16. Jännitän uusia tilanteita. Kadotan helposti itseluottamuksen.
24. I have many fears. I am easily scared.	24. Kärsin monista peloista, olen helposti pelästyvä.

Conduct problems scale	
5. I get very angry and often lose my temper.	5. Saatan vihastua kovasti ja menetän usein malttini.
7. I usually do as I am told.	7. Yleensä teen niinkuin minua käsketään.
12. I fight a lot. I can make other people do what I want.	12. Tappelen usein. Saan muut tekemään mitä haluan.
18. I am often accused of lying or cheating.	18. Minua syytetään usein valehtelemisesta tai petkuttamisesta.
22. I take things that are not mine from home, school or elsewhere.	22. Otan tavaroita, jotka eivät minulle kuulu, joko kotoa, koulusta tai muualta.
Peer problem scale	
6. I am usually on my own. I generally play alone or keep to myself.	6. Olen enimmäkseen yksinäni. Yleensä leikin yksin tai pitäydyn oman itseni seurassa.
11. I have one good friend or more.	*11. Minulla on yksi tai useampia hyviä ystäviä.*
14. Other people my age generally like me.	*14. Ikäiseni yleensä pitävät minusta.*
19. Other children or young people pick on me or bully me.	19. Olen muiden lasteen tai nuorten silmätikku tai kiusaamisen kohde.
23. I get on better with adults than with people my own age.	23. Tulen paremmin toimeen aikuisten kuin ikäisteni kanssa.

* positively worded problem items are in italics

6.3.4 The Dieting behaviour scale (publication III)

The school-survey questionnaire included questions on eating behaviour disorder items derived from DSM-IV (publication III). A short 9-item screening scale for assessing attitudes and behaviours involving dieting and weight concerns that could be easily answered by male and female adolescents was designed. The items were: "I want to be thinner"; "I exercise in order not to gain weight"; "I have been dieting"; "I am fearful of becoming fat"; "I have lost weight during a short time because I have not been eating enough"; "I am dissatisfied with my body"; "I am scared if my weight increases even a little"; "I can't control my eating"; "I practise self- induced vomiting after eating". The items were scored 0 (not true), 1 (somewhat true) and 2 (certainly true). The sum of items was used to generate an eating disorder total score which ranged from 0 to 18. A high level of dieting and weight concerns was defined as scoring at or above the 90th percentile of the total eating disturbance score. The internal consistency between items was acceptable (Cronbach's alpha 0.78). On the basis of self-reported height and weight, respondents were categorized in each sex and grade group as underweight (weight for height <15 percentile), normal weight (15-85 percentile) and overweight (>85 percentile).

6.3.5 The alcohol-use scale (publication IV)

The school survey included two items about alcohol use. The frequency of alcohol use was asked about with the question: "How often do you use alcohol (for example light beer)?" The drunkenness-orientation was asked about with the question: "How often do you use alcohol so that you get drunk?". The alternatives for answers in both items were:

"not at all", "once a month or more often", "once a week or more often". A high level of alcohol use was defined as getting drunk at least once a week. The family composition item was dichotomized to an intact two-parent family vs. other. The emotional and behavioural problems were assessed using the adolescent self-report version of the Strengths and Difficulties Questionnaire (SDQ-Fin) (Goodman et al. 1998). Concerning items about alcohol, 48 questionnaires were discarded because they were incomplete or inappropriately filled in, so altogether 1440 questionnaires were included in the statistical analysis. In items about alcohol the response rate was 87.7 % being 1.1 % lower than in other items included in the statistical analysis. In addition, background information was asked for: sex, age, grade, height and weight, the family composition (two biological parents vs. other) (publication IV).

6.4 Statistical methods

In publication I, the differences in sex, age and informants of the SDQ total difficulty scores and the sub-scores were analyzed with repeated measures analysis of variance. The inter-rater agreement between parent, teacher and self-report of the SDQ total difficulty scores and sub- scores, as well as the concurrent validity between the SDQ and the CBCL/YSR were analyzed with Pearson's correlation coefficient. The internal consistency of different SDQ scales was analyzed with Cronbach's alpha coefficient. The association between help-seeking variables and explanatory variables was analyzed with univariate logistic regression analysis. For explanatory variables, odds ratios (OR) and 95% confidence intervals (95% CI) were calculated. A p-value of 0.05 was interpreted as significant. The statistical analyses were carried out using the SAS system for Windows, release 6.12/1996.

In publication II, the SDQ self-reported total difficulties scores and the sub-scores, and the means and standard deviations grouped by sex and grade were analyzed by two-way analysis of variance, and the differences between the percentages of boys and girls at or above the 90th percentile cut-off points (at or below 10th percentile on the prosocial scale) were analyzed with the chi-square test. The correlations of the SDQ self-report items to the total difficulties score and to the sub-scores were analyzed with the Spearman correlation coefficient. The internal consistency reliability of different SDQ self-report items within each subscale was analyzed with Cronbach's alpha coefficient. The factor analyses of the SDQ self-report items were conducted using Varimax Orthogonal Transformation, with number of factors forced to be five (Nfactor criterion), and with number of factors left unspecified (Proportion criterion). The statistical analyses were carried out using the SAS system for Windows, release 6.12/1996.

In publication III, concerning the associations between the SDQ results and the dieting and weight concerns among Finnish adolescents, the associations between categorical

variables were analysed with the chi-square test. The associations between the explanatory variables and the outcome variables were analysed with multivariate logistic regression analysis. The 90th percentile cut-off point of the scoring distribution in the present sample was used in the SDQ total problem scale to indicate a high level of symptoms, as recommended previously (Goodman et al. 1998; Goodman 1997). Odds ratios (OR) and 95 % confidence intervals (95 % CI) were calculated. A p-value less than 0.05 was interpreted as significant. The statistical analyses were carried out using the SAS system for Windows, release 6.12/1996.

In publication IV, concerning the associations between the SDQ results and self-reported alcohol use among the Finnish adolescents, the differences between the explanatory variables and outcome variables were analysed with the chi-square test. The explanatory variables with a significant (p < 0.05) difference were included in the multivariate logistic regression analysis. All variables were dichotomized. The 90th percentile cut-off points of the scoring distribution of the present sample were used in the SDQ scales to indicate a high level of symptom loading, as recommended previously (Goodman et al. 1998). Odds ratios (OR) and 95 % confidence intervals (95 % CI) were calculated. The statistical analyses were carried out using the SAS system for Windows, release 6.12/1996.

7. RESULTS

7.1 Reliability of the SDQ-Fin

7.1.1 Internal consistency of the SDQ-Fin scales by different informants (publication I)

The internal consistency of the SDQ-Fin total scores was satisfactory in all three informants' reports (the Cronbach's alpha was 0.71, while the range in different subscales was 0.63 - 0.86). The alpha coefficient of the SDQ scales was rather high in most of the scales, and satisfactory or acceptable in the others. It was over 0.7 in all teacher-reported scales and over 0.6 in all other scales, except the conduct problem scale. As raters, the teachers had the best internal consistency (the mean level of internal consistency in different subscales being 0.79). The mean of internal consistency of subscales was 0.65 in self-reports and 0.67 in parent reports. The latter informants had internal consistencies of subscales in the same range, except for hyperactivity-inattention, which was more consistent in parent reports. When the subscales were compared, the lowest level of alpha was in the conduct and peer problems subscales (Table 3).

Table 3. Internal consistency reliability of the parent, teacher and self-report SDQ-Fin total scores and sub-scores analyzed with Cronbach's coefficient alpha.

Scores	self-report	parent report	teacher report
Total	0.71	0.71	0.71
Hyperactivity	0.66	0.73	0.85
Emotional symptoms	0.69	0.69	0.79
Conduct problems	0.57	0.59	0.72
Peer problems	0.63	0.64	0.73
Prosocial scale	0.69	0.68	0.86
Mean of subscores	0.65	0.67	0.79

7.1.2 Internal consistency of the SDQ-Fin self-report scales (publication II)

The internal consistency alpha of the SDQ-Fin self-reported total difficulties score, which includes all the problem scales (the prosocial scale was excluded), was 0.64 for all subjects, 0.69 for boys and 0.62 for girls. Of the subscales, the alpha was adequate on the emotional symptoms (0.71), hyperactivity (0.69), prosocial (0.68) and conduct problems scales (0.62). For the peer problems scale, the alpha was lower (0.53) (publication II).

7.1.3 Inter-rater agreement (publication I)

When the inter-rater agreement was studied between the SDQ-Fin total scores of the self-, parent and teacher reports (Table 4), it was moderate (0.38 - 0.44). The Pearson's correlation of the SDQ-Fin total difficulty scores between self-reports and parent reports

was 0.40, between self-reports and teacher reports 0.38, and between parent and teacher reports 0.44. The highest inter-correlation (r = 0.45) was on the hyperactivity scale between the ratings of parents and teachers. The lowest inter-correlation (r = 0.25) was on the emotional symptoms scale between self-reports and teacher reports.

Table 4. Inter-rater agreement: correlations between parent, teacher and self-report of SDQ total scores and subscores using Pearson's correlation coefficient.

Scores	self-report /parent report	self-report /teacher report	parent report /teacher report
Total	0.40*	0.38	0.44
Hyperactivity	0.39	0.34	0.45
Emotional symptoms	0.28	0.25	0.33
Conduct problems	0.28	0.30	0.30
Peer problems	0.39	0.38	0.39
Prosocial scale	0.37	0.28	0.29

*All the correlations were significant at level p < 0.001

Overall, if one looks at the results of inter-rater agreement, they show that the agreement between parents and teachers was highest, while between parent- and self-reports they are somewhat lower, and they are lowest between teacher and self-reports (publication I). The inter-rater agreement was high on the SDQ total difficulties and the hyperactivity-inattention scale by all pairs of informants, and it was high also on the peer problems by parent/self and by teacher/self agreements. The inter-rater agreement was rather low on the conduct problems scale (by parent/teacher and parent/self agreements), on the emotional symptoms (by parent/self and teacher/self agreements) and on the prosocial scale (by parent/teacher and teacher/self agreements).

7.1.4 Correlations of the SDQ-Fin self-report items (publication II)

The correlations of the SDQ-Fin self-report items with their respective sub-scales ranged from moderate to high (r = 0.47 - 0.73). The correlation of each item was highest with its respective SDQ-Fin sub-scale. The item correlations were highest on the emotional symptoms scale (r = 0.63 - 0.73), and somewhat lower on the hyperactivity (r = 0.59 - 0.72), on the prosocial behaviour (r = 0.58 - 0.70), and on the conduct problem scales (r = 0.55 - 0.66). They were lowest on the peer problem scale (r = 0.47 - 0.65). Of all items, "I am easily distracted. I find it difficult to concentrate" had the highest correlation to the total difficulties score. The correlations were calculated by the Spearman correlation coefficient (publication II).

7.1.5 Results of factor analyses (publication II)

The factor loading of each item with its respective sub-scale was calculated, and the program was forced to give a five-factor orthogonal solution (Varimax Orthogonal Transformation, Nfactor criterion), for boys and girls separately. For girls, the first three factors, the emotional symptoms, the prosocial behaviour and the conduct problems factor, were structured according to the SDQ-Fin subscales. All the respective items had their highest loadings on these factors (Table 9). Only three items out of a total of 25 items did not have their highest loading on the respective sub-scale. The fourth factor for girls was hyperactivity; three of the five respective items had the highest loadings on it. The fifth factor was peer problems; four of the five respective items had their highest loadings on this factor. The only exception was the item "Other people generally like me" which had the highest, although negative loading (- 0.37) on the prosocial factor. All five factors together explained 58.9 % of the variance.

For boys, the first three of the five specified factors were prosocial behaviour, emotional symptoms and hyperactivity factors. All the respective items had the highest positive loadings on these factors (Table 9). The fourth factor was peer problems and four of the five respective items had their highest loading on this factor (publication II).

The fifth factor for boys, the conduct problems, was somewhat problematic. Only one of the five respective items had its highest factor loading (0.81) on this factor ("I fight a lot. I can make people do what I want"). The conduct problem item "I get very angry and often lose my temper" had the highest factor loadings on the emotional symptoms (0.33) and hyperactivity (0.32) scales. The item "I usually do as I am told" had the highest loading on the hyperactivity scale (0.23). The item "I am often accused of lying and cheating" and the item "I take things that are not mine from home, school or elsewhere" had their highest loadings on the emotional symptoms scale (0.38 and 0.33, respectively). All five factors for boys explained 57.4 % of the variance.

7.2   Validity of the SDQ-Fin

7.2.1 Problematic behaviour and child mental help seeking according to parent (publication I)

The association between sex, grade and functioning within clinical range (>90th percentile) in the SDQ-Fin parent, teacher and self-reports, and having problematic behaviour according to parent (148/665, 22 %), was studied with univariate logistic regression analysis (Table 5). The results show that scoring at or above the 90th percentile (clinical range) was strongly associated with problematic behaviour according to parent in parent (OR 10.0), teacher (OR 3.2) and self-reports (OR 4.4) (publication I).

The parents were also asked if they had used or considered using mental health services because of their children's problems (publication I). Unfortunately, only 60% of parents (419/703) answered this item. Of these children, 16 had used, and in 10 cases, their parents had considered using mental health services for the children (6.2 % of the sample). When analyzed with univariate logistic regression analysis, the clinical range in the parent-report SDQ-Fin total difficulties scores of these 26 children was associated with service use or consideration to use them (OR 8.7, 95% CI 3.7-20.7, p<0.001).

Table 5. Associations between sex, grade, SDQ-Fin total scores of parent, teacher and self-reports and children/adolescents having minor or definite problems according to parent report. Analyses were performed with univariate logistic regression analysis.

variable*		n	%	OR	95% CI	p
sex						0.002
	girls	326	17	1.0		
	boys	339	27	1.8	(1.2 - 2.6)	
grade						0.008
	7, 9	284	17	1.0		
	1, 3, 5	381	26	1.7	(1.1 - 2.5)	
SDQ self-report						0.001
	normal range	427	20	1.0		
	clinical range	42	52	4.4	(2.3 - 8.4)	
SDQ parent-report						0.001
	normal range	598	18	1.0		
	clinical range	52	69	10.0	(5.3 - 18.6)	
SDQ teacher-report						0.004
	normal range	299	24	1.0		
	clinical range	28	50	3.2	(1.5 - 7.1)	

* n , % = number and percentage of parents reporting minor or definite problems; OR = odds ratio; 95% CI = 95% confidence interval

7.2.2 Concurrent validity

7.2.2.1 Correlations between the parent-rated SDQ-Fin and the CBCL (publication I)

The concurrent validity was assessed comparing the total difficulty scores and the sub-scores for the corresponding domains of the SDQ-Fin and the YSR/ CBCL with Pearson's correlation coefficient (Table 6). The correlation of the SDQ-Fin self-report and the YSR total scores for all subjects was 0.71; for boys 0.67 and for girls 0.76. The correlation of the parent SDQ-Fin and the CBCL total scores for all subjects was 0.75; for boys 0.62 and for girls 0.91 (all correlations were significant at the level p< 0.001) (publication I).

Table 6. Correlations between the Strengths and Difficulties Questionnaire (SDQ-Fin) parent reports and the Child Behaviour Checklist (CBCL), and between the SDQ-Fin self-reports and the Youth Self Report (YSR): total difficulties scores and the corresponding sub-scores.

Scores	Parent report	Self-report
	SDQ/ CBCL	SDQ/ YSR
Total/ Total	0.75	0.71
Conduct problems/ Externalizing	0.70	0.68
Conduct problems/ Delinquency	0.60	0.60
Conduct problems/ Aggressive	0.69	0.64
Hyperactivity/ Inattention	0.67	0.59
Emotional symptoms/ Internalizing	0.44	0.68
Emotional symptoms/ Withdrawn	0.34	0.43
Emotional symptoms/ Somatic problems	0.40	0.58
Emotional symptoms/ Anxious-depressed	0.44	0.68
Peer problems/ Social problems	0.41	0.51

* All correlations were significant at level p < 0.001

For the two parent reports, the correlations between the sub-scores were in the range of 0.34 - 0.70. The highest correlations were between the SDQ-Fin hyperactivity-inattention and the CBCL inattention, as well as between the SDQ-Fin conduct problems and the CBCL externalizing sub-scores; the latter was the result of the high correlation between the SDQ-Fin conduct problem scores and the CBCL aggression scores included in the externalizing scale. For the SDQ-Fin emotional symptoms, the highest correlation was with the CBCL anxious-depressed scores included in the internalizing scale (Table 6).

7.2.2.2   Correlations between the self-rated SDQ-Fin and the YSR (publication I)

For the two self-reports (Table 6), the correlations between the corresponding subscores were in the range of 0.43 - 0.68. The highest correlations were between the SDQ-Fin conduct problems and the YSR externalizing subscales, as well as between the SDQ-Fin emotional symptoms and the YSR internalizing subscales. The SDQ-Fin emotional symptoms had the highest correlation with the YSR anxious-depressed scores included in the internalizing scale. For both sets of questionnaires, the correlations were lowest between the SDQ-Fin emotional symptoms and the YSR withdrawn scores (publication I).

7.3   Descriptive properties of the SDQ-Fin

7.3.1 *The means of the SDQ-Fin parent, teacher and self-rated total difficulties scores (publication I)*

The means and standard deviations of the SDQ-Fin total difficulty scores rated by parents, teachers and youths themselves in Laitila-Pyhäranta sample are shown in Table

7. The comparisons were made between parent and self-reports, between parent and teacher reports and between teacher and self-reports with the repeated measures analysis of variance. The grades included in the analyses vary because the self-reports were not obtained from children in the first grade and the teachers in the secondary school did not take part in the study. In comparison of the parent and self-reports the effect of age was also analyzed by grouping the grades into primary (grades 3, 5) and secondary (grades 7, 9) school.

Table 7. The means and standard deviations of the SDQ-Fin total difficulties scores rated by children and adolescents, parents and teachers.

Grades	primary school		secondary school	
Sex	**boys** **mean (sd)**	**girls** **mean (sd)**	**boys** **mean (sd)**	**girls** **mean (sd)**
Self-report (n=528)	9.6 (5.4)	8.3 (6.0)	7.5 (5.1)	7.0 (4.5)
Parent report (n=703)	6.7 (4.4)	6.2 (4.7)	5.3 (3.9)	5.3 (4.9)
Teacher report (n=376)	7.8 (6.3)	4.8 (5.5)		

When the results of the parent and self-reports (grades 3, 5, 7 and 9 included) were compared, the effects of grade and informant were independently significant ($p < 0.05$) indicating that the children and adolescents rated higher total difficulty scores than their parents, and that the primary school children had higher total difficulty scores than the adolescents in the secondary school. There were no effects of sex or interaction. When the results of teacher and self-reports (grades 3, 5 included) were compared, the effect of sex was significant ($p < 0.001$). Boys had higher total difficulty scores than girls by both informants. There were no significant informant or interaction effects.

When the parent and teacher reports of the primary school children were compared, the effects of sex, informant and sex x informant interaction were significant ($p < 0.001$). The teachers rated higher SDQ-Fin total difficulty scores for boys and lower total difficulty scores for girls than the parents. It is worth mentioning that we looked further into reasons for this by examining the sub-scores of the parent and teacher reports. Significant differences were found in three sub-scales. The teachers were more likely to notice hyperactivity-inattention ($p < 0.01$) and conduct problems ($p < 0.05$) in boys than the parents. However, the results for the prosocial behaviour were converse. The teachers were more likely than the parents to rate girls functioning higher on prosocial behaviour than boys ($p < 0.05$). No significant differences were found on the emotional symptoms and peer problems scores.

7.3.2 The means of the SDQ-Fin self-report total difficulties scores and subscores (publication II)

The mean of the SDQ-Fin self-report total difficulties score in Salo-Rovaniemi sample was 11.2 (sd 5.4) for all subjects, 10.8 (sd 5.4) for boys and 11.7 (sd 5.3) for girls in the 7th and the 9th grades. Girls had significantly higher scores than boys in the total difficulties score ($p < 0.001$), on the prosocial behaviour scale ($p < 0.001$), on the hyperactivity scale ($p < 0.05$) and on the emotional symptoms scale ($p < 0.001$). Boys had significantly higher scores on the conduct problems scale ($p < 0.01$) and on the peer problems scale ($p < 0.05$). A significant age difference was found on the emotional symptoms scale, indicating that 9th graders had higher scores than 7th graders. A sex x age interaction ($p < 0.001$) was found on the prosocial behaviour scale, indicating that girls had higher scores on the prosocial scale than boys, and the difference was more significant among 9th graders than among 7th graders (Table 8).

Table 8. Means and standard deviations of the SDQ-Fin self-report sub-scores and total difficulties score rated by Finnish adolescents (n = 1458).

Grades	7th grade		9th grade	
Sex	boys mean (sd)	girls mean (sd)	boys mean (sd)	girls mean (sd)
The SDQ scores:				
Prosocial behaviour	6.1 (1.9)	7.1 (1.6)	5.9 (2.0)	7.5 (1.7)
Hyperactivity	3.4 (2.0)	3.6 (1.9)	3.5 (2.2)	3.8 (2.2)
Emotional symptoms	2.2 (1.8)	3.4 (2.2)	2.4 (2.0)	3.6 (2.3)
Conduct problems	2.5 (1.8)	2.2 (1.6)	2.6 (1.9)	2.3 (1.7)
Peer problems	2.5 (1.6)	2.4 (1.6)	2.5 (1.8)	2.2 (1.6)
Total Difficulties	10.5 (5.0)	11.6 (5.1)	11.0 (5.8)	11.8 (5.4)

7.3.3 The SDQ-Fin self-reported symptoms response rates (publication II)

In their SDQ-Fin self-report responses, the adolescents had a tendency to choose the middle alternative, i.e. they often had "somewhat" problems (Table 9). This frequently gave them a score of 1 and increased the sum of total scores. This tendency was emphasized on the positively worded problem items, four of which were rated "somewhat true" by 60 - 70 % of the adolescents. Emotional and hyperactivity problems were more frequent than conduct and peer problems. The severe problems rated ("certainly true") by more than 10 % of the adolescents were on the hyperactivity-inattention scale: restless (11.1 %), easily distracted (16.5 %), in addition to the item "I think before I do things" which 10.1 % of the adolescents rated "not true"; the emotional symptoms scale: "lot of headaches" (10.9 %), "worries" (11.2 %), "nervous in new situations" (17.4 %).

Table 9. Response rates of the SDQ-Fin self-report items and factor loadings for boys and girls, rated by Finnish adolescents (n = 1458).

	Response rates			Factor loadings*	
	Not true (%)	Somewhat true (%)	Certainly true (%)	Boys	Girls
Prosocial scale					
I try to be nice to other people. I care about their feelings.	2.3	45.9	51.8	0.61	0.51
I usually share with others (food, games, pens)	10.0	54.6	35.4	0.30	0.30
I am helpful if someone is hurt, upset or feeling ill.	3.8	54.2	42.1	0.62	0.58
I am kind to younger children.	6.0	42.9	51.0	0.50	0.39
I often volunteer to help others (parents, teachers, children).	12.1	69.3	18.7	0.63	0.57
Hyperactivity scale					
I am restless, I cannot stay still for long.	36.2	52.7	11.1	0.58	0.84
I am constantly fidgeting or squirming.	60.3	34.1	5.6	0.70	0.51
I am easily distracted. I find it difficult to concentrate.	30.2	53.4	16.5	0.46	0.31
*I think before I do things.***	10.1	59.3	30.6	0.33	0.28
I finish the work I'm doing. My attention is good	7.3	59.4	33.4	0.30	0.26
Emotional symptoms scale					
I get a lot of headaches, stomach-aches or sickness.	50.7	38.4	10.9	0.33	0.42
I worry a lot.	41.9	46.9	11.2	0.58	0.53
I am often unhappy, down-hearted or tearful.	68.4	26.3	5.3	0.46	0.52
I am nervous in new situations. I easily lose confidence.	31.7	50.9	17.4	0.42	0.56
I have many fears. I am easily scared.	68.6	26.2	5.1	0.53	0.64
Conduct problems scale					
I get very angry and often lose my temper.	41.5	49.9	8.6	0.22	0.45
I usually do as I am told.	5.5	69.1	25.4	0.04	0.22
I fight a lot. I can make other people do what I want.	72.7	23.9	3.4	0.81	0.61
I am often accused of lying or cheating.	72.0	21.3	6.7	0.16	0.45
I take things that are not mine from home, school or elsewhere.	75.9	20.6	3.5	0.21	0.42
Peer problem scale					
I am usually on my own. I generally play alone or keep to myself.	57.6	34.4	8.0	0.45	0.50
I have one good friend or more.	2.8	14.0	83.3	0.49	0.36
Other people my age generally like me.	5.3	66.2	28.5	0.38	0.34
Other children or young people pick on me or bully me.	73.1	23.8	3.1	0.37	0.31
I get on better with adults than with people my own age.	46.0	46.1	7.9	0.29	0.44

* Varimax Orthogonal Transformation. Specified five factor solution, ** *positively worded problem items in italics*

7.3.4 The Finnish SDQ self-report cut-off points (publication II)

The cut-off points of the SDQ-Fin total difficulties scores and the subscores were calculated for all subjects, as well as for boys and girls separately (Table 10). At or below the 10th percentile cut-off point on the prosocial scale covered significantly more boys (p < 0.001). The percentage of boys scoring at or above the 90th percentile cut-off point was significantly higher on the peer problem scale (p < 0.05). The percentage of girls scoring at or above the 90th percentile cut-off point was significantly higher on the emotional symptoms scale (p < 0.001) (publication II).

Table 10. Cut-off points at or above the 90th percentile of the SDQ-Fin self-report total difficulties score and the sub-scores rated by Finnish adolescents, and the percentage of boys and girls scoring at or above the 90th percentile (n = 1458).

	All subjects	Boys	Girls
Hyperactivity	6 (8.2 %)	6 (8.4 %)	6 (8.1 %)
Emotional symptoms	5 (12.1 %)	5 (6.4 %)	5 (17.8 %)
Conduct problems	4 (12.4 %)	4 (13.7 %)	4 (11.2 %)
Peer problems	4 (9.8 %)	4 (11.3 %)	4 (8.3 %)
Total difficulties	18 (9.4 %)	18 (9.4 %)	18 (10.6 %)
Prosocial behaviour*	4 (10.8 %)	4 (17.9 %)	4 (3.7 %)

* On the prosocial subscale, the cut-off point of the 10th percentile was used, and the percentage shows boys and girls scoring at or below the 10th percentile

7.4 Associations between the SDQ-Fin results and psychosocial problems among adolescents

The aim of the present study was to present the descriptive properties and factors associated with the self-reported eating disturbance symptoms and alcohol use of Finnish adolescents. The aim was also to explore further the associations between the emotional and behavioural problems described by the SDQ-Fin and, firstly, the eating disturbance symptoms and, secondly, the alcohol use of adolescents (publications III and IV).

7.4.1 *Associations between the SDQ-Fin and weight and dieting concerns (publication III)*

The percentage response rates of the self-reported eating disturbance items were calculated. The majority of girls in the 7th and 9th grades reported frequent weight control concerns. When the scores for the response alternatives "sometimes" and "often" are added together, 79 % of the girls and 44 % of the boys were dissatisfied with their current body shape. Of the girls, 76 % and of the boys, 61 % had "sometimes" or "often" exercised to avoid weight gain. Of the girls, 74 % and of the boys 37 % "sometimes" or "often" wanted to be thinner. Furthermore, of the girls, 7 % and of the boys, 3 % had practiced self-induced vomiting. Highly significant differences were observed between males and females in all the items. (Table 11)

Table 11. Percentual response rates of the self-reported eating disturbance items among female and male adolescents*.

Variable	Female (n = 725)		Male (n = 733)		p
	% sometimes	% often	% sometimes	% often	
Want to be thinner	37	37	28	9	<0.001
Exercise to avoid a gain in weight	56	20	41	20	<0.001
Dieting	24	17	11	3	<0.001
Fearful of becoming fat	34	30	17	6	<0.001
Lost weight	17	7	12	3	<0.001
Dissatisfaction with body shape	52	27	33	11	<0.001
Scared of weight increase	30	16	10	2	<0.001
Can't control eating	38	15	24	9	<0.001
Self-induced vomiting	4	3	1	2	<0.003

* Analyses were performed with the chi square test.

7.4.1.1 Factors associated with weight and dieting concerns

The 90th percentile cut-off point of the scoring distribution on the total eating disturbance scale was used to indicate a high level of dieting concerns. The associations between sex, grade, body mass, behavioural ratings, alcohol use and family composition, and extreme dieting concerns were studied using multivariate logistic regression analysis (Table 12). The results show that the strongest association with extreme dieting concerns was female sex (OR 10.4). Also SDQ-Fin total problem score (OR 4.6), alcohol use (OR 3.7) and body mass were independently associated with a high level of dieting concerns. Adolescents with normal weight or overweight had 3 to 4 times more extreme dieting concerns than those who were underweight. However, no significant difference was found between normal weight and overweight adolescents (Publication III).

To find the association between a high level of dieting concerns and the SDQ-Fin subscales, multivariate logistic regression analysis was performed. When the effect of

sex was controlled in the model, significant associations were found with emotional symptoms (OR 1.3 for one point rise, 95 % CI 1.2-1.4, p<0.001), conduct problems (OR 1.2 for one point rise, 95 % CI 1.1-1.4, p=0.005) and hyperactivity (OR 1.2 for one point rise, 95 % CI 1.1-1.3, p=0.007).

Table 12. Associations between explanatory variables and a high level of self-induced dieting behaviour and dieting concerns*.

	High level of eating distress		Multivariate analysis		
	n	%	OR	95%CI	p
Sex					<0.001
boys	719	2	1.0		
girls	725	17	10.4	5.6-19.2	
Body mass					<0.001
<15%	202	4	1.0		
15-85%	956	10	3.1	1.4- 7.2	
>85%	203	10	3.9	1.5-10.1	
Drunk weekly or more					<0.001
no	1329	9	1.0		
yes	102	23	3.7	1.9-7.1	
Grade					0.606
7th	718	10	1.0		
9th	726	9	0.9	0.8-1.2	
Family					0.459
two-parent	1007	9	1.0		
other	424	12	0.8	0.5-1.3	
Strengths and Difficulties Questionnaire					
Total scores					<0.001
normal	1266	7	1.0		
clinical**	132	30	4.6	2.7-7.8	

* Analysis was performed with multivariate logistic regression analysis; ** at or above the 90th percentile.

7.4.2 Associations between the SDQ-Fin and self-reported high alcohol use (publication IV)

7.4.2.1 Rates of alcohol use

About 5 % of the 7th graders and 14 % of the 9th graders reported that they used alcohol at least once a week. Furthermore, altogether 7% of the adolescents (100/ 1440 subjects), 4 % of the 7th graders, and 10 % of the 9th graders, reported that they got drunk at least once a week, which was defined as a high level of alcohol use (Table 13). About 60 % of the 7th graders, and 28 % of the 9th graders, reported that they never used alcohol. No gender differences were found in subjects with a high level of alcohol use.

Table 13. Descriptive data on the differences in alcohol use among Finnish 7th and 9th graders. Statistical analysis was performed with the chi-square test.

Explanatory variables		Total number of subjects (n= 1440)	Percentage of subjects with a high level of alcohol use	p-value
		n	%	
Sex				ns
girls		713	5.9	
boys		727	8.0	
Grade				<0.001
7th		720	3.8	
9th		720	10.1	
City				ns
Salo		647	7.6	
Rovaniemi		793	6.4	
Family composition				<0.001
intact two-parent family		1007	4.9	
other		421	11.4	
SDQ scales				
hyperactivity	< 90	1312	5.4	<0.001
	> 90	117	24.8	
emotional	< 90	1259	5.9	<0.001
	> 90	171	15.2	
conduct	< 90	1244	4.4	<0.001
	> 90	178	24.7	
peer problems	< 90	1282	6.9	ns
	> 90	140	7.1	
prosocial	> 10	1277	5.6	<0.001
	< 10	155	18.1	

< 90 = below 90th percentile; > 90 = at or above 90th percentile; < 10 = at or below 10th percentile; > 10 = abowe 10th percentile; ns = not significant

7.4.2.2 Associations with alcohol use

The chi-square comparison between explanatory variables and a high level of alcohol use was calculated. Furthermore, a high level of alcohol use was associated with other than an intact two-parent family, as well as with high levels of problems in hyperactivity, emotional and conduct behaviour, and a low level of positive prosocial behaviour. About 25 % of those with a high level of conduct and hyperactivity problems reported a high level of alcohol use (publication IV).

All the variables with a significant difference in the chi-square comparison were included in the multivariate analysis (Table 14). The conduct problems had the strongest association with a high level of alcohol use. Other variables independently associated with alcohol use were family composition, hyperactivity, emotional problems, a low level of prosocial behaviour and older age.

Table 14. The associations of several factors and a high level of alcohol use, i.e. drunk weekly (n = 100) analysed with multivariate logistic regression analysis.

Variable	OR	95 % CI
Conduct problems	4.0	2.4 - 6.9
Family composition other than intact two-parent-family	2.3	1.5 - 3.7
Hyperactivity	2.2	1.2 - 4.0
Emotional problems	1.9	1.03 - 3.4
Low level of prosocial functioning	1.9	1.03 - 3.4
9th grade	1.4	1.1 - 1.8

8.   DISCUSSION

8.1   Discussion of results

8.1.1 Reliability of the SDQ-Fin

The reliability of the SDQ-Fin was studied by exploring the internal consistency and inter-rater agreement (between parent-, teacher- and self-reported scores) of the SDQ-Fin scales (publication I), and by studying the SDQ self-report scale structure with item correlation analysis and factor analysis (publication II).

8.1.1.1 The internal consistency of the SDQ-Fin scales (publication I)

The internal consistency of the SDQ-Fin total difficulties in the parent, teacher and self-reports were shown to be satisfactory. As informants, the teachers showed the highest reliability in their ratings compared to the parent- and self-ratings. Parents and youths showed internal consistencies of the sub-scales in the same range, except for hyperactivity-inattention, which was more reliable in the parent reports (Table 3). The reason for teachers to be more reliable informants than parents may be because they are educationally a more homogeneous group and could be considered more reliable in understanding and answering written questions than the parents who are from all social classes and educational backgrounds (Kresanov et al. 1998).

For the SDQ-Fin parent report and teacher report total difficulties, the internal consistency of the SDQ-Fin was somewhat lower compared to results of previous studies (Appendix 1, Tables 15-17). However, for the SDQ-Fin sub-scales it was similar to that of other studies (Smedje et al. 1999; Goodman 2001; Malmberg et al. 2003; Muris et al. 2003; Wiedenfelt et al. 2003; Woerner et al. 2004). For example, the internal consistency of the parent-rated SDQ-Fin in the present study and in the Swedish study (Smedje et al.1999) was within the same range, the hyperactivity scale having the best and the conduct problems and peer problems scales having the lowest internal consistency in both studies (Appendix 1, Table 15).

For the SDQ sub-scales, the internal consistency of the hyperactivity-inattention problems has been shown to be highest in parent and teacher reports in previous studies (Goodman et al. 1998; Goodman 2001; Wiedenfelt et al. 2003; Muris et al. 2003). However, the youths themselves did not report their hyperactivity-inattention problems more reliably than their emotional symptoms, indicating that the youths are not as consistent in reporting their problems in concentration as parents and teachers (Appendix 1, Table 17).

Across all informants, the internal consistencies have been lowest for the conduct and peer problems scales in previous studies (Appendix 1, Tables 15-17). This weakness

may indicate that one or two items do not fit with the rest of the scale and/or problems are only partly reported. For the SDQ-Fin parent report, the reversely scored conduct problem and peer problem items ("I usually do as I am told", "has at least one good friend") and the low frequency conduct problem item ("steals") had a slightly negative influence on the internal reliability. The same conduct problem items also had the lowest intra-scale correlations in the SDQ-Fin teacher reports. Similar results were also found in other studies (Smedje et al. 1999; Muris et al. 2003).

For the SDQ-Fin self-report, the reversely scored hyperactivity and conduct problem items ("I finish the work I'm doing. My attention is good", "I usually do as I am told") had a slightly negative influence on the internal reliability. The latter conduct problem item also had the lowest intra-scale correlation with other conduct problem items ("steals" and "I get very angry and often lose my temper").

Somewhat variable psychometric properties of the SDQ self-report were also found by Rönning et al. (2004). They found that the intra-scale correlations were very low between oppositely posed questions (e.g. between item "I usually do as I am told" and other conduct problem items). They suggested that changing the formulation of these questions from positive to negative, as well as avoiding two sentences in the same item (e.g. "I fight a lot. I can make other people do what I want"), would increase the internal consistency of the scales, at least in the Norwegian version. However, changing the wording of the SDQ in the Norwegian version would make it difficult to compare Norwegian results with international studies.

8.1.1.2 The inter-rater agreement (publication I)

The reliability of the SDQ-Fin scales judged by the inter-rater agreement between the parent, teacher and self-reports was moderate. The agreement was higher between parents and teachers compared to parent/self ratings, while teacher/self ratings had the lowest agreement. However, the correlations in the present study were higher than in the meta-analytic study by Achenbach et al. (1987) which presented the results of 119 studies and included all types of informants; parents, teachers, mental health workers, observers, peers and subjects themselves. In their study, the mean correlation for the parent/teacher reports was 0.27, for the parent/self reports 0.25, and for teacher/self reports 0.20. The trend was similar in studies by Goodman 1997; Goodman et al. 1998; Goodman 2001; Wiedenfelt et al. 2003; Muris et al. 2003 (Appendix 1, Tables 18 - 20).

For the SDQ-Fin sub-scales, the results showed that the inter-rater agreement was highest, across all informants, for the hyperactivity-inattention problems, while the agreement was lowest for the prosocial behaviour. Agreement was also low between teachers and youths themselves for the emotional problems. The results were similar to those of previous studies (Goodman 1997; Goodman et al. 1998; Goodman 2001; Wiedenfelt et al. 2003;

Muris et al. 2003). Parents at home and teachers at school are better able to notice from their children's behaviour that they have problems in concentration than, for example, their emotional problems. However, children and adolescents may, understandably, tell more about their worries and sadness to their parents than to their teachers. However, already at school age, children start to spend more time outside the family. Parents and teachers may be less aware of what kind of problems children have with their friends.

8.1.1.3 Factor analysis of the SDQ-Fin self-report (publication II)

The postulated structure of the SDQ self-report was confirmed by the item correlation analysis in the present study. The correlation of each item was highest with its respective sub-scale (see Results 7.1.4.). According to the results of factor analyses for girls, the first three factors and all but 3 / 25 items were arranged according to the postulated five subscales. The exceptional items were two hyperactivity-inattention items (I am easily distracted ; I finish the work I'm doing) and one peer problem item (other people generally like me).

For boys, the factors generally confirmed the structure of four sub-scales (Table 9), except for one item on the peer problem scale, "bullied". However, the conduct problem factor was problematic. For boys, it associated with emotional symptoms and hyperactivity. Especially the item "I fight a lot. I can make people do what I want" seemed to have a different meaning for Finnish boys and girls. For boys it associated with feeling, e.g. restless, easily distracted, nervous and depressive.

Previously, the factor structure of the SDQ self-report has been studied, e.g., by Goodman (2001). The analyses showed that nearly all items loaded primarily, and usually exclusively on the predicted five factors in their study. The predicted five-factor structure fitted particularly well for the parent SDQ. For teacher and self-report SDQs, there was a tendency for any positively worded item to load on the prosocial factor.

In a Dutch study (Muris et al. 2003), for the parent-rated SDQ, the expected five factor structure emerged and all items loaded convincingly on the predicted factors. Only one prosocial behaviour item (considerate) had substantial secondary loading on the conduct problem factor (Appendix 3). Factor analysis of the self-rated SDQ showed a similar pattern, except that three conduct problem items (lies, tempers, obedient) and one peer problem item (good friend) did not load on their respective subscales.

The factorial structure of the SDQ self-report was also studied by Rønning et al. (2004). Their results showed that the confirmatory factor analysis generally affirmed the five-factor model and this model was similar for boys and for girls. However, the model of fit analysis indicated a somewhat variable and questionable fit. They also found some low

intra-scale correlations on the conduct problem, peer problem scales and hyperactivity-inattention, as well as low internal consistency on the conduct problem scale.

In the present study, when the results did not entirely confirm the presupposed structure of the SDQ scale, factor analysis was then conducted in a more exploratory way (with number of factors left unspecified), and three factors emerged (publication II). The first factor was primarily defined by three hyperactivity and two conduct problem items, indicating that these scales were more closely associated with each other than with other scales. The second factor consisted of the prosocial behaviour items, indicating that this was clearly a separate scale. The third factor was defined by three emotional symptom items and two peer problem items. These three factors may be interpreted as reflecting internalizing and externalizing dimensions in addition to prosocial behaviour.

Dickey and Blumberg (2004) found similar results. The results of their factor analysis of the SDQ parent report showed that the SDQ parent reports were based on three separate, but correlated, underlying dimensions, namely, prosocial behaviour and internalizing and externalizing problems (Appendix 3). These findings may indicate that this three-factor structure may be invariant across translations of the SDQ questionnaire, across different informants and rotation strategies (Dickey and Blumberg 2004).

The results of the factor analyses in the present study also give reasons to discuss the comorbidity of hyperactivity and conduct disorders, which have been found to be considerable among boys (McArdle et al. 1995; Sandberg 1996). The results of the present study may reflect genuine comorbidity in the sample if we assume that the hyperactivity and conduct problem scales of the SDQ self-report are independent and validly measure the postulated behavioural problems. For the SDQ parent report, Goodman and Scott's (1999) research results showed that the conduct problem and emotional symptom scales were independent of each other and, furthermore, the hyperactivity scale had good criterion and content validity, with the items covering the core symptoms in the diagnostic criteria of ADHD.

Further evidence for this kind of conclusion was found by Goodman (2001); there was very little overlap between the SDQ scales reflecting internalizing (emotional symptoms) and externalizing (hyperactivity-inattention and conduct problems) dimensions and, furthermore, the correlation between internalizing and externalizing was significantly lower for the SDQ than for the CBCL (Goodman & Scott 1999). However, the findings by Dickey and Blumberg (2004) were the opposite. When they further explored the factor structure of the parent-rated SDQ, they found that when the factor rotation method allowed the factors to correlate, externalization and internalization dimensions also emerged, and the overlap between internalizing and externalizing dimensions was closer in magnitude to correlations reported in studies using the longer questionnaire CBCL

(Achenbach 1991). This contradiction also gives reason for future research concerning the SDQ-Fin.

8.1.2 *Validity of the SDQ-Fin*

The aims of the present study were to explore the validity of the SDQ-Fin by studying the associations between mental health help-seeking because of a child's problematic behaviour according to the parent and the clinical range (at or above 90th percentile) by the SDQ-Fin (publication I), and by studying the concurrent validity by comparing the scores obtained, firstly with the SDQ-Fin parent report and the CBCL, and secondly with the SDQ-Fin self-report and the YSR (publication I).

8.1.2.1 Child mental help-seeking according to parent (publication I)

Referral to mental health services, service contact or services wanted for a mental health reason, or special education for a child's emotional and behavioural problems, has been used as one criterion with which to validate screening instruments (Bourdon et al. 2005). The actual referral usually means that the problems of the child have been persistent enough to make parents seek help (Achenbach and Edelbrock 1981).

In the present study, we asked the parents if they had used or had considered using the mental health services because of their child's problems (publication I). The SDQ-Fin total difficulties scores at or above the 90th percentile clinical range were strongly associated with these help-seeking variables in parent, teacher and self-reports. This supported the idea that the SDQ-Fin scores may be able to differentiate normal from clinical cases in a community sample. These findings are consistent with many reports (e.g. Goodman et al. 1998; Goodman and Scott 1999; Klasen et al. 2000; Muris et al. 2003; Wiedenfelt et al. 2003; Becker 2004 b) showing that the SDQ self-reported total difficulty score discriminated between a community sample and a psychiatric sample.

In the present study, the percentage of those parents who reported the use of or their intention to use mental health services for their children was small (6 %), including the 4 % of parents who reported that they had used the child mental health services because of their child's problems. However, there has been a significant increase in use of services since then. When use and need of child mental health services was studied in a Finnish community sample, the use of services increased from 4 % in 1989 to 12 % in 2005 among eight-year-old boys. The corresponding figures for girls of the same age were 1 % and 4 %. The result indicated that the threshold for seeking help has become lower. Furthermore, in 2005, of the screen-positive boys, 32 % had been referred, whereas the corresponding figure for girls was only 15 % (Sourander et al. 2008). A low rate of mental health service use has been found among those children who are at risk of a psychiatric disturbance or who meet the criteria for psychiatric symptoms (Leaf et al.

1996; Offord et al. 1987; Verhulst and Ende 1997; Zahner et al. 1992; Sourander et al. 2004). Especially the mental health needs of girls are poorly recognized (Sourander et al. 2008).

8.1.2.2 The concurrent validity (publication I)

The concurrent validity was measured by comparing the total scores and the sub-scores of the parent- and self-report SDQ-Fin and the Achenbach questionnaires (the CBCL and the YSR). The correlations between the total scores generated by the parent reports (0.75) and self-reports (0.71) in the two instruments were satisfactory, given the great difference in length and in item contents of the two measures. The correlations in the present study were in the same range as in previous studies by Wiedenfelt et al.(2003) and Muris et al. (2003), but somewhat lower than Goodman and Scott (1999) found in their study in Britain, where the parent-rated SDQ and the CBCL correlated highly with one another, the correlation between the total difficulties scores being 0.87. In Germany, Klasen et al. (2000) found that the correlations of the total scores between the parent-rated SDQ/ the CBCL and the self-rated SDQ/ the YSR were 0.78 and 0.77, respectively, thus being intermediate between the Finnish and the British results (Appendix 1, Tables 21 - 22).

Concerning the concurrent validity between the sub-scores in the two parent reports, the correlations were higher between the externalizing dimensions, while the correlations between the internalizing dimensions were lower, especially for the SDQ-Fin emotional problems and the CBCL withdrawn behaviour (Appendix 1, Table 21). The difference in the content between the two questionnaires seems more prominent for the internalizing dimension. However, for the two self-reports (the SDQ and YSR), the correlations between the internalizing and externalizing dimensions were in the same range, although the correlation between the SDQ-Fin emotional symptoms and the YSR withdrawn scale were the lowest of all subscales (Appendix 1, Table 22).

In the present study, we were not able to compare the two questionnaires by their ability to differentiate between children drawn from community and clinical samples. However, it is worth mentioning that Goodman and Scott (1999) found that the parent-rated SDQ and the CBCL were equally able to discriminate between children drawn from high-risk (psychiatric) and low-risk samples. Furthermore, the interview-based ratings correlated more highly with the SDQ than with the CBCL scores, with this difference being statistically significant for inattention-hyperactivity. Other results were along similar lines (Klasen et al. 2000; Goodman 2001; Muris et al. 2003; Wiedenfelt et al. 2003). However, the composite internalizing problem score of the CBCL was better able to detect emotional disorders than the five-item emotional symptom scale of the parent SDQ, while the SDQ parent and teacher reports were superior in discriminating between patients with or without hyperkinetic/attention deficit disorder (Becker et al. 2004 a).

8.1.3 Descriptive properties of the SDQ-Fin

The aim of the present study was to present the descriptive properties of emotional and behavioural problems of Finnish children and adolescents measured by the SDQ-Fin, that is, the means, symptom response rates and cut-off points (publications I and II).

8.1.3.1 The parent-, teacher-and self-rated means of the SDQ-Fin total difficulties (publication I)

The parents in Laitila-Pyhäranta sample reported fewer problems when describing their 9 - 16 -year-old children compared to the youngsters' self-reports as measured by the SDQ-Fin total difficulties. This was consistent with other studies, when parent reports and children's and adolescents' self-reports have been compared. The children and adolescents (11 – 19-year-olds) report more emotional and behavioural problems than their parents reported about them (Seiffge-Krenke and Kollmar 1998; Stanger and Lewis 1993; Verhulst and van der Ende 1997). All informants in the present study reported fewer total difficulties than found, e.g. in Britain (www.sdqinfo.com) (Appendix 1, Table 23).

The younger children (9-11 years old) had more problems than the adolescents (13-16 years old) when rated by the parents and the youngsters themselves in the present study. This was also in accordance with many previous results (Seiffe-Krenke and Kollmar 1998; Verhulst et al. 1985).

In addition, some more current results are along the same lines. Muris et al. (2003) found that in the SDQ self-reports this age effect was primarily due to peer problems of which younger children had more than older ones. Parents reported younger children to have more hyperactivity problems than older ones in a study by Woerner et al. (2004). When they compared the results of the SDQ parent and self-reports, no differences between boys and girls were found. This result was similar in the present study. This is contrary to other reports, where the parent-reported SDQ total difficulties scores have been found to be higher for boys (Smedje et al.2001; Muris et al. 2003; Wiedenfelt et al. 2003; Woerner et al. 2004) (publication I).

The teachers reported boys (9 – 11 years old) to have more and girls to have fewer behavioural problems than the parents reported. The teachers were more likely to notice especially externalizing problems (hyperactivity and conduct problems) in boys and less likely to notice them in girls than the parents. However, the results for the prosocial behaviour were the opposite, the teachers were more likely than parents to notice prosocial behaviour in girls than in boys. Although the results of the previous studies vary concerning the differences in the behavioural problems of boys and girls, this is in accordance with some studies. For example, in four out of five comparisons of four different nationalities, teachers scored boys higher than girls in the TRF total problem scores (Verhulst and Achenbach 1995).

8.1.3.2 The means of the SDQ-Fin self-report (publication II)

In the present study, adolescent girls (the 7th and 9th graders in the Salo-Rovaniemi sample) reported more problems according to the SDQ total difficulties scores than boys of the same age. This was due to the hyperactivity-inattention problems and emotional symptoms. Girls also had a better level of prosocial behaviour, which increased with age for girls and decreased with age for boys. Boys had significantly more conduct problems and peer problems. The present results raise many interesting questions for future research. For example, do girls have more inattentive than hyperactivity type of symptoms, and does their hyperactivity associate with other psychological problems (e.g. depression)?

The self-reported total difficulties in the present study (publication II) were comparable with findings in the South-Norwegian study by Van Roy et al. (2006). The results of their large-scale community study (n = 29631, age range 10-19 years) showed that in one age group (older adolescents, 16-20 years), girls reported more problems than boys, although boys reported more problems than girls in preadolescence (10-13 years). Furthermore, Norwegian youths, both boys and girls, reported more hyperactivity-inattention problems than youths in Finland, North-Norway and the U.K (Koskelainen et al.; Rønning et al.; www.sdqinfo.com) (Appendix 1, Table 24). Finnish girls reported more problems, on average, than was found in a recent British study (www.sdqinfo.com), and in another Norwegian study by Rønning et al. (2004), where they did not find gender differences for the SDQ total difficulties, and boys had more hyperactivity-inattention problems than girls. For the other sub-scales, their results were similar to those of the present study (Appendix 1, Table 24).

8.1.3.3 The SDQ-Fin self-reported symptoms response rate (publication II)

Symptom response rates of the SDQ-Fin self-reported emotional and behavioural problems (publication II) demonstrate that the adolescents (in the 7th and 9th grades) had a tendency to choose the middle alternative ("somewhat problems") when rating the items (Table 9). This tendency may, at least partly, explain the often found difference in the amount of emotional and behavioural problems according to the adolescents' self-reports compared to their parents' reports (Stanger & Lewis 1993; Verhulst & Van der Ende 1992). The parents' tendency to rate their children as having "no problems" was found in the Swedish study of the SDQ parent reports of their 6 - 10 -year-old children (Smedje et al. 1999). Furthermore, in their study the SDQ prosocial items were rated "certainly true" by 48 - 90 % of the parents, compared to 19 - 52 % of the adolescents in the present study. These results indicate that there may be wide differences in various informants' ways to respond to the questionnaire items.

In the present study, the self-reported hyperactivity and emotional problems were more frequent than conduct and peer problems (Table 9). The severe problems rated by more than 10 % of Finnish adolescents ("certainly true" in problem items or "not true" in positively worded items) were on hyperactivity-inattention: being restless; easily distracted and having difficulties to concentrate; in addition to the item "I think before I do things" (reflective) and on emotional symptoms: somatic problems (headaches, stomach-aches, sickness); worries; nervousness in new situations (clingy). All problems (except one: reflective) experienced as severe by more than 10 % of Finnish adolescents were included among the severe problems reported by more than 10 % of Norwegian boys and girls (Van Roy et al. 2006).

Van Roy et al. (2006) found that 19 – 22 % of Norwegian 13 to 16-year-old adolescents, both boys and girls, rated severe problems on hyperactivity items (except 11 % of boys and 8 % of girls on reflective-item). More than 10 % of adolescents reported severe problems also on emotional items (somatic; worries; clingy) as well as on one conduct problem item (getting very angry and losing temper) and on one peer problem item (being usually on ones own, playing alone and keeping to oneself).

8.1.3.4 The SDQ-Fin cut-off points (publication II)

The cut-off points of the SDQ-Fin peer problems, emotional problems and prosocial behaviour covered a significantly different percentage of boys and girls (in the 7th and 9th grades). The abnormal range (at or above the 90th percentile) was the same as the British (www.sdqinfo.com) cut-off points for three subscales, namely, emotional symptoms, conduct problems and hyperactivity-inattention. Compared with the Norwegian studies (Van Roy et al. 2006; Ronning et al. 2004), the Finnish cut-offs were generally one point lower (Appendix 1, Table 25).

The percentile scores may be used, e.g. if the researcher wants to know whether a sample of girls with high SDQ scores is comparable to age- and sex-matched controls. However, originally, Goodman proposed the cut-off points of the SDQ scores to act as useful "warning signals" for the clinician that the adolescent may have a disorder and be in need of some preventive or clinical action (Goodman 1997; Goodman et al. 1998).

Discussing the procedure of choosing the optimal cut-off point, Malmberg et al. (2003) found that the 90th percentile cut-off score was not the same as the optimal cut-off score of the SDQ total difficulties scale according to ROC analysis (Appendix 2). Still, it is often assumed that the true prevalence of psychiatric disorders at any given time in a community sample would normally not lie above 10 % (Achenbach and Edelbrock 1981; Ford et al. 2003). Consequently, the 90th percentile in a community sample is often chosen as the threshold for different child psychiatric screening instruments (e.g.Goodman 1997; Almquist et al. 1988). A useful strategy for researchers may be to choose cut-offs according to the likely disorder rate in the target sample, and according to the relative importance for

the study of false positives and false negatives (Goodman 1997). However, the study by Malmberg et al. (2003), as well as other previous studies (Kresanov et al. 1998; Fombonne 1991), indicate that the use of ROC analyses based on comparisons between clinical and non-clinical samples are valuable complements to traditional significance tests.

8.1.4 *Associations between the SDQ-Fin results and psychosocial problems*

8.1.4.1 Associations between the SDQ-Fin self-report results and dieting and weight concerns (publication III)

When the descriptive properties of eating disturbance symptoms were studied, it was found that they were extremely common among Finnish adolescents, indicating that, especially among girls, they are almost normative concerns. About three-fourths of the girls reported that they were sometimes or often dissatisfied with their body, wanted to be thinner and exercised in order not to gain weight, while two-thirds had fears of gaining weight. There were no age differences between 7th and 9th graders in high level of eating distress.

The results were similar to the findings of previous studies (e.g. Casper and Offer 1990; Story et al. 1991; Grigg et al. 1996; Patton et al. 1999). The results of a more recent international survey conducted in 35 countries and regions, including Finland, concerning health behaviour of 11-15-year-olds (Currie et al. 2004) also showed clear gender differences. Although the percentages in their study were somewhat lower than in the present study, more girls (36 %) than boys (22 %) reported dissatisfaction with their body weight (for example, feeling a bit too fat or much too fat), and more girls (18 %) than boys (8 %) reported dieting and weight control behaviour. However, in their study, girls' dissatisfaction and dieting behaviour increased with age, while it remained about the same level among boys in different age groups.

Female adolescents seem to be more critical toward their body than male adolescents. The bodily changes and the increase in body fat that accompanies puberty in females have been found to be a consistent risk factor for eating disturbances (Fombonne 1997). Patton et al. (1999) found that female adolescents who dieted at a severe level were 18 times more likely to develop a new eating disorder within six months than those who did not diet. However, given that only a small minority of the adolescents with a high level of dieting concerns may in fact develop eating disorders, additional factors must operate that cause such adolescents to remain preoccupied with their weight and thinness, while others outgrow the preoccupations.

In the present study, however, the percentage of boys who reported that they were often dissatisfied with their body shape and had often exercised to avoid gaining weight, was also rather high (11 % and 20 % respectively). Similar to our results, Childress et al. (1991) reported that boys had more concerns about weight than had been found in previous studies;

of the 9- to 16-year-old normal-weight boys, 28 % reported that they "felt they looked fat", and 13 % were "afraid to eat because of weight gain". However, boys may sometimes give different psychological meaning than girls to bodily exercise and weight loss. Paxton et al. (1991) found that nearly two-thirds of girls and boys believed that being thinner would have an impact on their lives, but the majority of girls believed the impact would be positive, while the majority of boys believed it would be negative.

The majority of adolescents, in the present study, with a high level of self-induced dietary restrictions were of normal-weight or overweight. These results were similar to those found in previous studies that both normal weight and overweight youths are likely to express weight-specific concerns and engage in self-induced dieting behaviours (Patton et al. 1997). However, for the overweight adolescents, the dieting behaviour may have a different, i.e. a more health promoting meaning than for the normal-weight adolescents. Overweight of adolescents has become an increasing problem for school-aged children and for school health professionals. Contrary to body dissatisfaction and dieting behaviour results, in a number of countries, including Finland, boys have higher rates of overweight than girls (Currie et al. 2004; Luopa et al. 2006).

When the associations between psychological problems measured by the SDQ-Fin and eating disturbance symptoms were studied, it was found, as expected, that a high level of emotional symptoms was associated with a high level of weight and dieting concerns. This was consistent with previous results that adolescents' preoccupation with weight and dieting are likely to indicate psychological problems (Caspar and Offer 1990). Casper and Offer (1990) found that increased dieting was associated with greater body- and self-image dissatisfaction, with depressed mood, and overall symptomatic distress. Unexpectedly, externalizing symptoms, that is, conduct problems and hyperactivity, were also associated with dieting behaviour in the present study. However, this is in accordance with results by Thompson et al. (1999), who found that eating disturbance was significantly associated with aggressive behaviour (for example, fighting, battery and assault) in adolescence (publication III).

A frequent high level of alcohol use was also associated with a high level of dieting concerns in the present study, as was expected. Research on the relationship between problem drinking and disordered eating among adolescents has produced conflicting results. Studies of community-based populations have shown that the frequency of dieting in early adolescence predicts later alcohol use in adolescence (Krahn et al. 1996). It has been suggested that body-image concerns and dieting play a central role in the association of alcohol problems and disordered eating among adolescent girls.

Female dissatisfaction with body weight and appearance are commonly explained by the social and cultural ideal of thinness which is associated with femininity. Female adolescents try to control their changing body more often than males through weight-

loss behaviours. The transition from childhood through adolescence to womanhood with psychological and bodily changes and with future possibilities of adult sexual relationships is a major developmental task. Heebink et al. (1995) found that young anorectic girls under 14 years of age had fewer depressive and anxiety feelings but, at the same time, more maturity fears than the older anorectic girls. These maturity fears may influence them to diet, lose weight, and avoid the physiological changes that belong to adolescent development.

8.1.4.2 Associations between the SDQ-Fin self-report results and high alcohol use (publication IV)

The present study showed a sharp increase in alcohol use from 7th to 9th grade. This was consistent with other studies conducted at about the same time (Karvonen 1995; Rimpelä et al.1996; Lintonen et al. 2000). In a more recent school-heath survey, 3 % of boys and 2 % of girls in the 8th grade, and 6 % of boys and 4 % of girls in the 9th grade, reported that they got drunk at least once a week. The percentages are somewhat lower than those found in the present study. The drunkenness oriented use of alcohol among adolescents has decreased in the last six years. The percentage of those who got drunk once a month, has decreased from 27 % to 17 % (Luopa et al. 2006).

In the present study, boys and girls reported equally often high level of alcohol use. Similar results were found in a recent European survey where equal proportions of Finnish boys and girls reported using alcohol and having been drunk (Hibell et al. 2003). Also Lintonen et al. (2000) found that gender differences have diminished, with girls approaching boys in alcohol use and drunkenness. An increase in alcohol consumption among girls has also been found in other studies (Persson et al. 1994; Silbereisen 1995; Bränström et al. 2007). In many countries, also the delinquency rates among girls have risen more sharply than those among boys (Smith 1995; Robins and McEvoy 1990) (publication IV).

In the present study, conduct problem behaviour was the most potent factor associated with alcohol use. Furthermore, hyperactivity problems increased the risk of alcohol use. This is in line with our assumptions and with previous studies. Conduct problems have been shown to be predictive of all forms of substance use and abuse (Robins and McEvoy 1990; Silbereisen et al.1995; Shoal et al. 2007). The combination of conduct disorder and hyperactivity carries a particularly high risk. Chilcoat et al. (1999) found that hyperactivity lowered the threshold for risk of substance use when associated with externalizing problems. Also children with attention-deficit problems have been found to be at significantly higher risk of a specific negative course characterized by antisocial and substance-related disorders (Mannuzza et al. 1998). Furthermore, mood disorder in combination with conduct disorder and substance abuse is a strong risk factor for teenage suicides (Schaffer et al. 1996).

According to the present study, the SDQ-Fin self-reported emotional problems and a low level of prosocial functioning were also associated with alcohol use. This result is in line with studies where depression has been found to be associated with alcohol use (Windle and Davies 1999). In depressed teenagers, the progression from alcohol or drug initiation to the onset of substance abuse has been found to occur more rapidly (Rao et al. 1999). King et al. (1996) found that depressed girls who used alcohol had longer depressive episodes, more conduct problems, psychosocial impairment and over-involvement with boys than depressed girls who did not use alcohol. For depressed boys, alcohol abuse was predictive of conduct disorder, older age and schoolwork problems.

8.1.4.3 Psychosocial problems: clinical implications

In addition to general screening of the emotional and behavioural problems of children and adolescents, there are many important clinical issues not covered by the SDQ. We studied two of them, namely, the self-reported weight and dieting concerns and high alcohol use of adolescents. These issues also present a major challenge for school health providers.

Because the concerns over weight and dieting are so common among adolescents, clinicians who assess or treat them should be aware that there is a high prevalence of dissatisfaction with body weight and self-induced dietary restriction already by the time of early adolescence. Although few adolescents will actually develop anorexia nervosa or bulimia, some may benefit from early intervention regarding healthier eating behaviours. It is important to be aware that a high level of weight and dieting concerns may also indicate a high level of emotional distress. Efforts are needed to provide appropriate guidance and education about normal changes during puberty. The present study suggests that education directed toward helping children and adolescents to adopt healthy eating and exercise habits should take place at an early age, already before the age of 13.

In the present study, a frequent high level of alcohol use was associated with a high level of dieting concerns. Research on the relationship between problem drinking and disordered eating among adolescents has produced conflicting results. Recent studies of community-based populations have shown that the frequency of dieting in early adolescence predicts later alcohol use in adolescence (Rao et al. 1999). Another challenge for school health providers is how to help adolescents to avoid early heavy alcohol use, especially if the youngsters also have conduct problems or hyperactivity, and to minimize their risk of becoming later abusers.

8.2   Discussion of methods

8.2.1 Strengths and limitations of the study

The present study can be considered important for many reasons. According to the author's knowledge, the SDQ is in clinical use, but other research results besides the present study have not been reported from Finland. The study also has practical importance, in examining the psychometric properties of methods which are in everyday use in child psychiatry. Major advantages of the study were also its three samples and their sizes, in addition to the good response rates. This issue is more thoroughly discussed in the next section. Another advantage of the present study was the possibility for cross-informant comparisons. We obtained the SDQ scores from parents, teachers and youths themselves (publication I). This enabled us to better understand some of the differences usually found in results by different informants.

However, there were also several limitations which should be taken into consideration when interpreting the results of the present study. Some limitations of this study are shared with other epidemiological studies based on questionnaires. Given the large number of subjects, the rating scale approach provides useful information but lacks the specificity of more narrowly directed questionnaires and, on the other hand, the additional depth that child psychiatric interviews might provide.

The study design was a cross-sectional school survey. Although in one of the samples, the data were obtained in an identifiable manner (publication I), we had no possibility to arrange a follow-up. We had results from different age groups but they were from different age cohorts, not continuous development followed over years. Furthermore, the likely clinical cases (that is, those children and adolescents who require further assessment and, possible treatment or other interventions) were defined from scoring distributions on the SDQ scales. Because of the lack of clinical interviews, it was not possible to validate them against clinically defined diagnoses.

In addition, we gathered information about some common psychosocial problems of children and adolescents, which are not covered by the SDQ, namely, weight and dieting concerns and high alcohol use, but many important issues remained outside the scope of the study; a range of more specific child psychiatric symptoms (including sleeping problems, enuresis, depression, phobias, obsessive and psychotic symptoms etc.). Also because of the methodology, we were not able to study other possible contributing family factors such as problems in family functioning or parental psychopathology. The only family variable used was family structure (two biological parents vs. other).

Although the overall response rate of the study was good, another weakness of this study is the lack of information about the characteristics of the parents, teachers and youths who did not participate. One whole primary school with 33 pupils in Laitila did not

participate and nine more children did not get their parents permission to participate, together accounting for 5.7 % of the target population (n = 735). It is possible that some of these were children and adolescents who were experiencing more problems than those who did participate.

8.2.2 Samples

The three samples of the study were gathered in a cross-sectional manner, using questionnaire screens in a community sample of school-aged children and adolescents. Although they were not random samples, they were large enough to allow generalizations. In the first sample from Laitila and Pyhäranta, the SDQ-Fin results were obtained from multiple informants and in several age groups in primary and secondary school (n = 703). The response rate in the first study was 96 % for the SDQ-Fin parent report, 92 % for the teacher report and 88 % for the self-report. As a limitation, all the participating children were living in a suburban/ rural area, which may have introduced a bias into the results (publication I). However, in a wide epidemiological study, no significant difference was found in emotional and behavioural problems of children between the different types of municipalities, i.e. city, suburban and rural Finland (Puura et al. 1995). Contrary to this, in a Swedish study (Larsson and Frisk 1999), children and adolescents living in large cities had more behavioural problems than those living in rural areas. In the present study, the means of the SDQ self-report total difficulties were different in the two samples, from Laitila-Pyhäranta (publication I) and from Salo-Rovaniemi (publication II). They were about 2-3 points higher in the latter sample, the participants of which were living in two towns. The urban/rural dimension may have affected the results.

The aim of the present study was also to examine the concurrent validity of the SDQ-Fin, i.e. to compare the parent and self-rated SDQ-Fin with the Achenbach questionnaires (CBCL and YSR). In the second sample from Laitila and Pyhäranta, the response rate was 94 % for the SDQ-Fin self-report (n = 129), but unfortunately, the response rate for the parent reports in this sample was somewhat lower (59 %) and the sample size was smaller (n = 81) (publication I). This may have had an effect on the results and may help to explain some of the inconsistencies we found. The correlation between the SDQ-Fin emotional symptoms and the CBCL Internalizing scores was lower for parent-rated (but not for self-rated) SDQ-Fin than in other comparable European studies.

The samples from Salo and Rovaniemi (publication II) were rather large but the age range was narrower than in the samples from Laitila and Pyhäranta. The response rate for the SDQ-Fin self-report (n = 1458) was 89 %. Another limitation was that the results were based only on adolescents' self-reports, and some adolescents may have belittled while others may have exaggerated their emotional and behavioural problems. However, there was also another difference in the samples from Laitila-Pyhäranta and Salo-Rovaniemi, namely, identifiability/anonymity. When the means of the SDQ-Fin self-report were

compared to other studies, they were close to the results in Norwegian studies by Van Roy et al. (2006) and Rønning et al. (2004) who also obtained their results anonymously. These results indicate that it is important to take into consideration the identifiability factor when interpreting results in this and future studies.

8.2.3 Methods

Several questionnaires were constructed by the researchers for the purposes of the present study. For the validation of the SDQ-Fin, help-seeking items examining the use of child mental health services according to parents were included in the school survey (publication I). Although the overall response rate of the study was good, unfortunately, the response rate was lower for the help-seeking question (60 %). Furthermore, the outcome measure, need for and use of child mental health services, relied solely on reports obtained from the parents; there was no verification of the service use external to parents' reports, which indicates that reported service use may have been underestimated. The help-seeking variable was used as a criterion to validate the clinical range of the obtained SDQ-Fin scores. In spite of the somewhat low response rate, the high SDQ-Fin scores correlated well with parent-reported service need and use, and the result was as expected.

An additional item was included about the parents' view of whether the child has significant emotional or behavioural difficulties as recommended by Goodman (1999) in the extended version of the SDQ: "Overall do you think that your child has difficulties in one or more of the following areas: emotions, concentration, behaviour or being able to get on with other people?" When using the SDQ extended version (Goodman 1999), it was a striking finding how well the presence or absence of psychiatric disorder could be predicted from this single item. Caseness predicted from this item was generally only slightly less accurate than caseness predicted from the combination of symptoms and impact.

The school-survey questionnaire also included items on weight and dieting concerns and alcohol use. The dieting items were derived from DSM-IV. A short 9-item screening scale for assessing attitudes and behaviours involving dieting and body shape that could be easily answered by male and female adolescents was designed. The internal consistency between items was acceptable (Cronbach's alpha 0.78). The school survey also included two items about alcohol use. In items about alcohol use the response rate was 87.7 %, being only 1.1 % lower than in other items included in the statistical analysis. In addition, background information was asked for: height and weight, and family composition (two biological parents vs. other).

8.2.3.1 Comments on the statistical methods used in the present study

The participants filled in questionnaire items by choosing from three alternatives which were coded from zero to two. The total difficulties score was 0 – 40, and sub-scale

scores were 0 – 10. The associations between categorical variables were analyzed with the chi-square test (e.g. see Table 9). Spearman's correlation coefficient for categorical variables was used when analyzing the correlation of each SDQ self-report item with the total difficulties score, and each item with the respective subscore (see Results, 7.1.4). Goodman (1997) used Pearson's correlation coefficient in his study but reported that the results were similar when using Spearman's correlation coefficient.

The present study also used parametric statistical methods which are based on normality assumptions, namely, Pearson's correlation coefficient, analysis of variance, and regression analysis. An important reason for using these parametric methods was to enable comparisons between the present results and the results of previous studies (e.g. Goodman 1994; 1997; 1998; 1999; Smedje et al. 1999; Klasen et al. 2000). In these parametric methods, the SDQ total difficulties score or the sub-scores were treated as normally distributed. This assumption can not be strictly valid, since the variables are discrete. However, especially the SDQ total difficulties score can be treated as normally distributed since it can take on any of 41 different values, and since it is calculated by summing up 20 different item scores. Indeed, visual inspection showed that the distribution of the total difficulties score within the groups by sex and grade looked normal. The SDQ sub-scores can take on only 11 different values, so they have fewer possible values, and their distributions were in many instances somewhat skewed.

Analysis of variance and linear regression analysis are based on the assumption that the error terms of the linear model are independent and normally distributed, and have the same variance. However, the normality assumption is not crucial: namely, the usual test procedures (t-tests and F-tests) produce approximately the right conclusions, under fairly general study conditions, when the sample size is large, provided that the error terms are independent and identically distributed (Arnold 1981). When the sample sizes are large, it is therefore customary to apply the usual procedures of linear regression analysis even when the errors are known not to be normally distributed. The study arrangements tried to create independence between the responses of different subjects. Teachers who gave the questionnaires in the classes, were all advised to do it in a standardized way and to emphasize that pupils should think for themselves.

In the present study, factor analysis was first used in a confirmatory way (Varimax orthogonal transformation with number of factors forced to be five) because the SDQ questionnaire has five different subscales. When the results did not entirely confirm the presupposed structure of the SDQ scale, factor analysis was conducted in a more exploratory way (with a number of factors left unspecified) and three factors emerged. The hyperactivity-inattention and conduct problems were combined into externalizing problems, emotional and peer problems were combined into internalizing problems, and the third factor was prosocial behaviour (publication II).

A special characteristic of the present study design was that the children were rated by multiple informants (parents, teachers and youths themselves) and the ratings were compared when calculating the inter-rater reliability. Usually this is calculated from observations made of the same material by different informants (e.g. when two researchers are independently rating the same videotape). However, in the present study, each informant has access only to a part of the whole information about the child, that is, the child or adolescent knows his or her own ideas and feelings better than his or her parents, who see how the child behaves at home but not as much of how the child behaves at school as teachers do. Therefore Goodman (1997) recommends using the term inter-rater agreement instead of inter-rater reliability for this kind of study design.

8.2.4 Comparison of some SDQ-Fin psychometric properties with international studies

Overall, the results of this study showed that the psychometric properties of the Finnish version of the SDQ were adequate, and for the most part supported its reliability and validity. The Finnish adolescents' self-reports contributed to the understanding of the discrepancies between reports of different informants. Furthermore, the postulated structure of the SDQ self-report was confirmed by the correlation analysis in the present study.

However, there were also some variable findings of the psychometric properties. The inter-rater agreement was, across all informants, lowest for prosocial behaviour among all the subscales. In their self-reports, adolescents had a tendency to choose the middle alternative ("somewhat problems") when rating the items. However, parents in community samples may have a tendency to rate their children as having "no problems". This was found in the Swedish study by Smedje et al. (1999). Furthermore, in their study, the SDQ prosocial items were rated "certainly true" by 48 - 90 % of the parents, compared to 19 - 52 % of the adolescents' ratings in the present study. These results explain some of the differences found in various informants' ways to respond to the questionnaire items.

In the present study, internal consistencies of the SDQ peer problem and conduct problem scales were rather low, indicating that the items included measure somewhat different properties. Reversely scored items and low frequency items had a slightly negative influence on the internal consistency. Rønning et al. (2004) also found very low intra-scale correlations between reversely posed conduct problem items. They suggested that changing the formulation of these questions from positive to negative, as well as avoiding two sentences in the same item, would increase the internal consistency of the scales, at least in the Norwegian version. Although it is important to study the reasons for these somewhat low internal reliabilities of the SDQ scales, the present version of the SDQ has already been translated into many languages and is frequently in use all over

the world. If a researcher wants to be able to compare results with international studies, no changes in the original wording can be recommended.

Furthermore, the results of the factor analyses showed that especially for boys, the conduct problems scale was problematic. Most of the conduct problem items associated with emotional symptoms and hyperactivity-inattention. When the factor analysis was conducted in an exploratory way, three factors reflecting internalizing and externalizing dimensions and prosocial behaviour emerged. These findings were similar to the findings of Dickey and Blumber (2004), and indicate that this kind of three factor-structure may be invariant across translations of the SDQ questionnaire, across different informants and factor rotation strategies.

8.2.5 *Clinical implications and aims of further research*

The results of the present study provide additional information on the usefulness of the SDQ-Fin as a screening instrument for epidemiological research and for clinical purposes. Currently the reliability and validity of the SDQ has been widely studied internationally, although there are not very many studies with large representative samples. However, large-scale epidemiological studies with the SDQ have been conducted, e.g. in Britain (Goodman 2001), in the USA (Dickey and Blomberg 2004; Bourdon et al. 2005) and in Norway (Van Roy et al. 2006). Also in Finland, the SDQ-Fin might be a suitable method for nationwide epidemiological research to screen the mental heath of 3 – 16-year-old children and adolescents.

The SDQ might be chosen, when the researcher or clinician wants to assess the mental health risks in groups of children and adolescents with a brief and not too time-consuming questionnaire. In clinical use, it might be a suitable screening instrument in primary health care or in school health care to promote early identification of a psychiatric disorder, that is, to find those children and adolescents who require further assessment and possible treatment or other interventions.

In special health care, the SDQ may not add much to the information on and understanding of children's problems on an individual examination level, but it may be very useful in the initial assessment, serving as a baseline of a child's emotional and behvioural problems for a follow up. When using the SDQ to detect the likely psychiatric cases, it is recommended to use at least two informants. When only one informant can be used, parents or teachers are a better choice than the SDQ self-report alone (Goodman 2003; Goodman et al. 2004).

In clinical practice, the SDQ may be supplemented with questionnaires directed at specific problems, clinical interview and observations. In combination with the SDQ it would be easy to add the semi-structured interview DAWBA (Goodman et al. 2000),

to the selection of methods. The SDQ is a short and general questionnaire; it does not cover a range of more specific child psychiatric symptoms (including sleeping problems, enuresis, depression, phobias, obsessive compulsive and psychotic symptoms).

The SDQ self-report cut-off point obtained in the present study for the total difficulties score was similar to that found in Norway (Rønning et al. 2004), but the cut-off points for the sub-scores were usually one point lower than those found in Norway (Rønning et al. 2004; Van Roy et al. 2006). However, the cut-off points of the present study may be recommended to be used among 13- to 17- year-old Finnish adolescents, tentatively and according to the aims of the study, until more representative community and clinical samples with wider age ranges are obtained.

A limitation of the present study was that we had no clinical samples to study the ability of the SDQ-Fin to differentiate clinical cases and child psychiatric disorders. However, many international studies of concurrent validity between the SDQ and independent clinical evaluations have shown that the SDQ has a clinical predictive ability (Goodman 2001; Goodman et al. 2003). Relying on these international results, the SDQ may be recommended for use also in clinical settings in Finland. Goodman has (1999) suggested that a combination of symptom and impact scores would be the best indicator of caseness. Although the personal impact of psychiatric symptoms is a part of contemporary diagnostic criteria, only few SDQ studies have used such information in their analyses (Rothenberger and Woerner 2004). In future, examining the reliability and validity of the SDQ-Fin should continue using more representative community and clinical samples. It would also be important to start research with under-school-aged children using the SDQ-Fin parent version for 3 – 4-year-olds.

Interesting questions for future research also gave the results in the present study that Finnish school-girls reported more problems in their self-reports than Finnish boys; this was due to emotional symptoms, but also to hyperactivity-inattention problems (publication II). Do girls have more inattentive than hyperactivity types of symptoms, and does their hyperactivity associate with other psychological problems (e.g. depression)? Is there a time trend when girls have as many or more problems than boys, and what might be a reason for such a trend? One might even speculate that changes in parents' methods of raising their children may have a changing effect on the future mental health problems of children and adolescents. Girls and boys growing today are tomorrow's parents of their children.

In the present study, girls also reported, more than boys, prosocial behaviour which increased with age for girls and decreased with age for boys (publication II). Furthermore, teachers were more likely than parents to notice prosocial behaviour in girls than in boys (publication I). The development of the prosocial behaviour in a child is a complicated process where awareness of norms, self-regulation, empathy, shame and guilt intertwine

(Hay 1994). In future research, assessment of prosocial factors and their associations with emotional and behavioural problems should be emhasized more. For example, could two groups of youths be found with a high level of problems, one group with a high and the other with a low level prosocial behaviour. For some children and adolescents prosocial behaviour may be a protective factor which may have a positive effect on their prognoses. This may be found in follow-up studies.

Follow-up of child psychiatric interventions will be important in future SDQ research. Along with the aim of developing evidence-based treatment interventions, there is also a need to develop methods to assess these interventions. This is recommended by many prestigious scientists, e.g. by Robert Goodman and Peter Fonagy. They are both members of the CAMHS Outcome Research Consortium (CORC), which collaborates the child and adolescent mental health services (CAMHS) across the U.K. Their aim is to encourage a common model of routine outcome evaluation. The CORC handbook (www. corc.uk.net) recommends the SDQ as one of the routinely used methods for evaluation of interventions in child and adolescent psychiatric clinical practice. For example, in a training course of psychotherapy (ESCAP 2007), chaired by Peter Fonagy, it was stated, that in their clinical practice, the aim of their intervention is, at least to be able to reduce the SDQ scores of the patient down from the highest 95th percentile.

8.2.6 Conclusion

Since the beginning of the SDQ research in Finland almost ten years ago, the use of and research on the SDQ has spread all over the world in a fascinating way. The fact that the SDQ is freely available on the internet made it easily accessible. There have been aspirations to establish norms and validate the SDQ in different cultures. However, in many countries, the samples studied have not been representative enough to span the full age range in both genders. Furthermore, comparing the present research results with results from various cultures is not an easy task because of the different samples, divided into different groups sometimes by age, sometimes by both age and gender etc. Some variability and questionability in the results of the psychometric properties of the SDQ have also been found, and suggestions have been made for improvement, e.g. in item wording, but changes, at least in the original English version, are not recommended because of the wide previous research. We do not know how minor differences in wording and meaning affect study participants' processing of the SDQ items. Collaborative efforts in international research of the SDQ are recommended. However, the SDQ has been proved to be useful in large-scale epidemiological research, in screening and in predicting clinical caseness at initial assessment. Future research may provide evidence of its sensitivity to change in follow-ups and in the evaluation of child psychiatric intervention outcomes.

ACKNOWLEDGEMENTS

Viivi Koistinen, my dearest daughter, who has given me the joy of life, who has helped me to keep both feet on the ground, and who has reminded me of what the most important things are in life.

REFERENCES

Achenbach TM (1991a) Manual for the Child Behavior Checklist/4–18 and 1991 Profile. University of Vermont, Department of Psychiatry, Burlington, VT.

Achenbach TM. (1991b) Manual for the Teacher's Report Form and 1991 Profile. Burlington: University of Vermont. Department of Psychiatry, VT.

Achenbach TM (1991c) Manual for the Youth Self-report and Profile. Burlington: University of Vermont, Department of Psychiatry, VT.

Achenbach TM, Edelbrock CS (1981) Behavioral problems and competencies reported by parents of normal and disturbed children aged 4 through 16. Monographs of the Society for Research in Child Development 46(188):1-82.

Achenbach TM, McConaughy SH, Howell CT (1987) Child / adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. Psychological Bulletin 101:213-232.

Almagram MH, Shuwail AY (2004) Validity of the self-report version of the strengths and difficulties questionnaire in Yemen. Saudi Medical Journal 25(5):592-601.

Almquist F, Bredenberg P, Suominen I, Leijala H. (1988) Social kompetens och beteendeproblem bland skolbarn och barnpsykiatriska patienter - en empirisk studie med CBCL. (Social competence and behaviour problems among school children and child psychiatric patients - an empirical study using CBCL). Nordisk Psykiatrisk Tidsskrift 42:311-319 (In Swedish).

Almqvist F, Tuompo-Johansson E, Panelius E, Aronen E, KairemoA-C (1991) Screening for psychiatric symptoms on the basis of the Rutter Parent's Questionnaire. In: Tamminen T, Almqvist F, Piha J (eds.) Psychiatric symptoms in Finnish children: an epidemiological study of 1000 8-year old children in five regions of the country. Reports of Psychiatrica Fennica 91.

Arnold SF (1981) The theory of linear models and multivariate analysis. John Wiley and Sons. USA:141–158.

Becker A, Woerner W, Hasselhorn M, Banasschewski T, Rothenberger A (2004a) Validation of the parent and teacher SDQ in a clinical sample. European Child and Adolescent Psychiatry 13:II/11-II/16 (Supplement 2).

Becker A, Hagenberg N, Roessner V, Woerner W, Rothenberger A (2004b) Evaluation of the self-reported SDQ in a clinical setting: Do self-ratings tell us more than ratings by adult informants? European Child and Adolescent Psychiatry 13:II/17-II/24 (Supplement 2).

Biederman J, Mick E, Faraone SV, Braaten E, Doyle A, Spencer T, Wilens TE, Frazier E, Johnson MA (2002) Influence of gender on attention deficit hyperactivity disorder in children referred to a psychiatric clinic. American Journal of Psychiatry 159:36-42.

Bourdon KH, Goodman R, Rae DS, Simpson G,Koretz DS (2005) The Strengths and Difficulties Questionnaire: U.S. Normative Data and Psychometric Properties. Journal of the American Academy of Child and Adolescent Psychiatry 44(6):557-564.

Braun DL, Sunday SR, Halmi KA (1994) Psychiatric comorbidity in patients with eating disorders. Psychological Medicine 24:859-867.

Conners CK (2001) Conners' Rating Scales – Revised. New York, Multi-Health Systems.

Currie C, Roberts C, Morgan A, Smith R, Settertobulte W, Sandal O, Barnekow Rasmussen V (eds.) (2004) Young People's Health in Context: international report from the HBSC 2001/02 survey. WHO Policy Series: Health policy for children and adolescents Issue 4, WHO Regional Office for Europe, Copenhagen.

Dickey WC, Blumberg SJ (2004) Revisiting the Factor Structure of the Strengths and Difficulties Questionnaire: United States, 2001. Journal of the American Academy of Child and Adolescent Psychiatry 43(9):1159-1167.

Elander J, Rutter M (1996) Use and development of the Rutter parents' and teachers' scales. International Journal of Methods in Psychiatric Research 6:63-78.

Fergusson DM, Hornwood LJ (1995) The prevalence and risk factors associated with abusive or hazardous alcohol consumption in 16-year-olds. Addiction 90:935-948.

Fillmore KM, Hartka E, Johnstone BM, Leino EV, Motoyoshi M, Temple MT (1991) The collaborative alcohol-related longitudinal project. A meta-analysis of life course variation in drinking. British Journal of Addiction 86:1221-1268.

Fombonne E (1997) Eating disorders: Time trends and possible explanatory mechanisms. In: Rutter M, Smith DJ, (eds.). Psychosocial disorders in young people. New York:Wiley & Sons: 616-685.

Fombonne E (1991) The Use of Questionnaires in Child Psychiatry Research: Measuring their Performance and Choosing an Optimal Cut-Off. Journal of Child Psychology and Psychiatry 32(4):677-693.

Ford T, Goodman R, Meltzer H (2003) The British child and adolescent mental health survey 1999: The prevalence of DSM-IV disorders. Journal of the American Academy of Child and Adolescent Psychiatry 42(10):1203-1211.

Garner DM, Olmsted MP, Bohr Y, Garfinkel PE (1982) The Eating Attitude Test: Psychometric properties and clinical correlates. Psycholgical Medicine 12:871-878.

Garner DM (1991) Eating Disorder Inventory 2, professional manual. Odessa, FL: Psychological Assessment Resources.

Goodman R (1994) A Modified Version of the Rutter Parent Questionnaire Including Extra Items on Children's Strengths: A Research Note. Journal of Child Psycholgy and Psychiatry 35: 1483–1494.

Goodman R (1997) The Strengths and Difficulties Questionnaire: A Research Note. Journal of Child Psychology and Psychiatry 38:581–586.

Goodman R (1999) The extended Version of the Strengths and Difficulties Questionnaire as a Guide to Child Psychiatric Caseness and Consequent Burden. Journal of Child Psychology and Psychiatry 40:791–799.

Goodman R (2001) Psychometric properties of the Strengths and Difficulties Questionnaire. Journal of the American Academy of Child and Adolescent Psychiatry 40(11):1337-1345.

Goodman R, Meltzer H, Bailey V (1998) The strengths and difficulties questionnaire: A pilot study on the validity of the self-report version. European Child and Adolescent Psychiatry 7:125 –130.

Goodman R, Scott S (1999) Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: Is small beautiful? Journal of Abnormal Child Psychology 7:17–24.

Goodman R, Refrew D, Mullick M (2000) Predicting type of psychiatric disorder from Strengths and Difficulties Questionnaire (SDQ) scores in child mental health clinics in London and Dhaka. European Child and Adolescent Psychiatry 9:129-134.

Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29-36.

Hawes DJ, Dadds MR (2004) Australian data psychometric properties of the Strengths and Difficulties Questionnaire. Australian and New Zealand Journal of Psychiatry 38:644-651.

Hay DF (1994) Prosocial development. Journal of Child Psychology and Psychiatry 35:29-71.

Holmila M (1995) Intoxication and hazardous use of alcohol: Results from the 1992 Finnish drinking habits study. Addiction 90:785-794.

Hudziak JJ, Derks EM, Althoff RR, Rettew DC, Boomsma DI (2005) The genetic and environmental contributions to attention deficit hyperactivity disorder as measured by Conners' Rating Scale – Revised. American Journal of Psychiatry, 162 (9):1614-20.

ICD-10 (1999) International Statistical Classification of Diseases and related Health Problems 1 (10th revision).

Kadesjo B, Janols LO, Korkman M, Mickelsson K, Strand G, Tillingsgaard A, Gillberg C (2004) The FTF (Five to Fifteen): the development of parent questionnaire for the assessment ADHD and comorbid conditions. European Child and Adolescent Psychiatry 13: 3-13 (Supplement 3).

Karvonen S (1995) Regional differences in drinking among Finnish adolescents. Addiction 90: 57-66.

Kashala E, Elgen I, Sommerfelt K, Tylleskar T (2005) Teacher ratings of mental health among school children in Kinshasa, Democratic Republic

of Congo. European Child and Adolescent Psychiatry 14:208-215.

Keski-Rahkonen A, Sihvola E, Raevuori A, Kaukoranta J, Bulik CM, Hoek HW, Rissanen A, Kaprio J (2006) Reliability of Self-Reported Eating Disorders: Optimizing Population Screening. International Journal of Eating Disorders 39:754-762.

Klasen H, Woerner W, Wolke D, Meyer R, Overmeyer S, Kasnitz W, Rothenberger A, Goodman R (2000) Comparing the German Versions of the Strengths and Difficulties Questionnaire (SDQ-Deu) and the Child Behaviour Checklist. European Child and Adolescent Psychiatry 9:271-276.

Koskelainen M, Sourander A, Helenius H (2000a) Dieting and weight concerns among Finnish adolescents. Nordic Journal of Psychiatry 55:427–431.

Koskelainen M, Sourander A, Kaljonen A (2000b) The Strengths and Difficulties Questionnaire among Finnish school-aged children and adolescents. European Child and Adolescent Psychiatry 9:277-284.

Koskelainen M, Sourander A (2001a) Self-reported alcohol use and behavioural problems among Finnish adolescents. Psychiatria Fennica 32:57–66.

Koskelainen M, Sourander A, Vauras M (2001b) Self reported Strengths and Difficulties in a community sample of Finnish adolescents. European Child and Adolescent Psychiatry 10:180-185.

Kovacs M (1992) Children's Depression Inventory, CDI. Manual. Ontario: Multi-Health Systems Inc.

Kresanov K, Tuominen J, Piha J, Almqvist F (1998) Validity of child psychiatric screening methods. European Child and Adolescent Psychiatry 7:85-95.

Larsson B, Frisk M (1999) Social competence and emotional/behaviour problems in 6-16 year-old Swedish school children. European Child and Adolescent Psychiatry 8:24-33.

Leaf PJ, Alegria M, Cohen P, Goodman SH, Horwitz S, Hoven CW, Narrow WE, Vaden-Kiernan M, Regier DA (1996) Mental health service use in the community and schools: results from the four-community MECA study. Journal of the American Academy of Child and Adolescent Psychiatry 35:889-897.

Lifshitz F, Moses N (1988). Nutritional dwarfing: growth, dieting and fear of obesity. Journal of the American College of Nutrition 7:367-376.

Luopa P, Pietikäinen M, Puusniekka R, Jokela J, Sinkkonen A (2006) Kouluterveyskysely 2006. Etelä-Suomen lääninraportti. Stakes.

Malmberg M, Rydell A-M, Smedje H (2003) Validity of the Swedish version of the Strengths and Difficulties Questionnaire (SDQ-Swe). Nordic Journal of Psychiatry 57(5):357-363.

Maloney MJ, McGuire J, Daniels SR, Specker B (1989) Dieting behavior and eating attitudes in children. Pediatrics 84:482-489.

Mallick MH (1983) Health hazards of obesity and weight control in children: a review of the literature. American Journal of Public Health 73:78-82.

McArdle P, O'Brien G, Kolvin I (1995) Hyperactivity: Prevalence and relationship with conduct disorder. Journal of Child Psychology and Psychiatry 36:279-303

McGee R, Williams S, Bradshaw J, Chapel JL, Robins A, Silva PA (1985) The Rutter scalefor completion by teachers: Factor structure and relationship with cognitive abilities and family adversity for a sample of New Zealand Children. Journal of Child Psychology and Psychiatry 26(5):727-739.

Marzocchi GM, Capron C, Di Pietro M, Tauleria ED, Duyme M, Frigerio A, Gaspar MF, Hamilton H, Pithon G, Simoes A, Therond C (2004) The use of Strengths and Difficulties Questionnaire (SDQ) in Southern European countries. European Child and Adolescent Psychiatry 13:II/40-II/46 (Supplement 2).

Mathai J, Andreson P, Bourne A (2004) Comparing psychiatric diagnoses generated by the Strengths and Difficulties Questionnaire with diagnoses made by clinicians. Australian and New Zealand Journal of Psychiatry 38:639-643.

Messer SC, Angold A, Costello EJ and Burns BJ (1996) The Child and Adolescent Burden Assessment (CABA): measuring the family impact of emotional and behavioral problems. International Journal of Methods in Psychiatric Research 6:261-284.

Mullick MSI, Goodman R (2001) Questionnaire screening for mental health problems in Bangladeshi children: a preliminary study. Social Psychiatry and Psychiatric Epidemiology 36: 94-99.

Muris P, Meesters C, van den Berg F (2003) The Strengths and Difficulties Questionnaire (SDQ). Further evidence for its reliability and validity in a community sample of Dutch children and adolescents. European Child and Adolescent Psychiatry 12:1-8.

Obel C, Heiervang E, Rodriguez A, Heyerdahl A, Smedje H, Sourander A, Gudmundsson OO, Clench-Aas J, Christensen E, Heian F, Mathiesen KS, Magnússon P, Njar∂vík U, Koskelainen M, Rønning JA, Stormark KM, Olsen J (2004) The Strengths and Difficulties Questionnaire in the Nordic countries. European Child and Adolescent Psychiatry 13:II/32 – II/39 (Supplement 2).

Offord DR, Boyle MH, Szatmari P, Rae-Grant NI, Links PS, Cadman DT, Byles JA, Crawford JW, Blum HM, Byrne C, Thomas H, Woodward CA (1987) Ontario child health study II. Six-month prevalence of disorder and rates of service utilization. Archieves of General Psychiatry 44:577–838.

Patton GC, Selzer R, Coffey C, Carlin JB, Wolfe R (1999) Onset of adolescent eating disorders: population based cohort study over 3 years. British Medical Journal 318:765-768.

Patton GC, Carlin JB, Shao Q, Hibbert ME, Rosier M, Selzer R, Bowes G (1997) Adolescent dieting: healthy weight control or borderline eating disorder? Journal of Child Psychology and Psychiatry 38:299-306.

Persson E, Hanson BS, Råstam A-S (1994) Alcohol habits among teenagers in Sweden: Factors of importance. Journal of Studies on Alcohol 55:719-725.

Puura K, Tamminen T, Bredenberg P, Escartin T, Kaukonen P, Leijala H, Rutanen M, Suominen T, Salmelin R (1995) Lasten psyykkiset häiriöt Tampereen yliopistollisen sairaalan vastuualueella. (The psychiatric disturbances of children in the Tampere university catchment area). Reports of Psychiatria Fennica (Supplement 110) (In Finnish).

Reynolds CR, Richmond BO (1978) What we think and feel: a revised measure of children's manifest anxiety. Journal of Abnormal Child Psychology 6:271-280.

Rimpelä M, Jokela J, Luopa P, Liinamo A, Huhtala H, Kosunen E, Rimpelä A, Siivola M (1996). School health 1996 -survey. (In Finnish). Helsinki: STAKES Aiheita 40.

Rohde P, Lewinsohn PM, Seeley JR (1996). Psychiatric comorbidity with problematic alcohol use in high school students. Journal of the American Academy of Child and Adolescent Psychiatry 35:101-109.

Rothenberger A, Woerner W (2004) Strengths and Difficulties Questionnaire (SDQ) – Evauations and applications. European Child and Adolescent Psychiatry 13:II/1-II/2 (Supplement 2).

Rutter M (1967) A children's behaviour questionnaire for completion by teachers: preliminary findings. Journal of Child Psychology and Psychiatry 8:1-11.

Rutter M, Graham P (1966) Psychiatric disorder in 10- and 11-year-old children. Proceedings of Royal Society of Medicine 59:382-387.

Rønning JA, Handegaard BH, Sourander A, Morch W-T (2004) The Strengths and Difficulties self-report Questionnaire as a screening instrument in Norwegian community samples. European Child Adolescent Psychiatry 13:II/73-II/82 (Supplement 2).

Rønning JA, Stormark KM (2004) The Strengths and Difficulties Questionnaire in the Nordic countries. European Child Adolescent Psychiatry 13:II/32-II/39 (Supplement 2).

Sandberg S (1996) Hyperkinetic or Attention Deficit Disorder. British Journal of Psychiatry 169: 10-17.

Samad L, Hollis C, Prince M, Goodman R (2005) Child and adolescent psychopathology in a developing country: testing the validity of the Strengths and Difficulties Questionnaire (Urdu version). International Journal of Methods in Psychiatric research 14(3):58-166.

Schoemaker C, Verbraak M, Breteler R, van der Staak C (1997) The discriminant validity of the Eating Disorder Inventory -2. British Journal of Clinical Psychology 36:627–629.

Schleimer K (1983) Dieting in teenage schoolgirls. Acta Paediatrica Scandinavica 72 (Supplement 312).

Seiffge-Krenke I, Kollmar F (1998) Discrepancies between mother' and fathers' perceptions of sons' and daughters' problem behaviour: a

longitudinal analysis of parent-adolescent agreemen on internalizing and externalizing problem behaviour. Journal of Child Psychology and Psychitary 39:687-697.

Smedje H, Broman J-E, Hetta J, von Knorring A-L (1999) Psychometric properties of a Swedish version of the "Strengths and Difficulties Questionnaire". European Child and Adolescent Psychiatry 8:63-70.

Sourander A, Multimäki P, Santalahti P, Parkkola K, Haavisto A, Helenius H, Nikolakaros G, Piha J, Tamminen T, Moilanen I, Kumpulainen K, Aronen ET, Linna S-L, Puura K, Almqvist F (2004) Mental health service use among 18-year-old adolescent boys: A prospective 10-year follow-up study. Journal of the American Academy of the Child and Adolescent Psychiatry 43(10):1250–1258.

Sourander A, Niemelä S, Santalahti P, Helenius H, Piha J (2008) Changes in psychiatric problems and service use among 8-year-old children: A 16-year population-based time-trend study. Journal of American Academy of Child and Adolescent Psychiatry 47(3):317-327.

Story M, Rosenwinkel K, Himes JH, Resnick M, Harris LJ, Blum RW (1991) Demographic and risk factors associated with chronic dieting in adolescents. American Journal of Diseases of Children 145:994-998.

Stanger C, Lewis M (1993) Agreement among parents, teachers and children in internalizing and externalizing behaviour problems. Journal of Clinical Child Psychiatry 33:107-115.

Taylor E, Schacher R, Thorley G and Wieselberg H (1986). Conduct disorder and hyperactivity: I. Separation of hyperactivity and antisocial conduct in British child psychiatric patients. British Journal of Psychiatry 149:760-767.

Thompson KM, Wonderlich SA, Crosby RD, Mitchell JE (1999) The neglected link between eating disturbances and aggressive behavior in girls. Journal of the American Academy of Child and Adolescent Psychiatry 38:1277-1283.

Van Roy B, Groholt B, Heyerdahl S, Clench-Aas J (2006) Self-reported strengths and difficulties in a large Norwegian population 10-19 years. Age and gender specific results of the extended SDQ-

questionnaire. European Child and Adolescent Psychiatry 15:189-198.

Verhulst FC, Achenbach TM (1995) Empirically based assessment and taxonomy of psychopathology: Cross-cultural applications. A review. European Child and Adolescent Psychiatry 4:61-76.

Verhulst FC, van der Ende J (1997) Factors associated with child mental health service use in the community. Journal of the American Academy of Child and Adolescent Psychiatry 36:901-909.

Verhulst FC, Akkerhuis GW, Althaus M. (1985) Mental health in Dutch children: (I) a cross-cultural comparison. Acta Psychiatrica Scandinavica 72 (Supplement 323).

Vostanis P (2006). Strengths and Difficulties Questionnaire: research and clinical applications. Current Opinion in Psychiatry 19(4):367-372.

Weir K, Duveen G (1981) Further development and validation of the prosocial behaviour questionnaire for use by teachers. Journal of Child Psychology and Psychiatry 22:357-374.

Wiedenfelt van BM, Goedhart AW, Treffers PDA, Goodman R (2003) Dutch version of the Strengths and Difficulties Questionnaire. European Child and Adolescent Psychiatry 12:281-289.

Woerner W, Becker A, Rothenberger A (2004) Normative data and scale properties of the German parent SDQ. European Child and Adolescent Psychiatry 13:II/3-II/10 (Supplement 2).

Woerner W, Fleitlich-Bilyk B, Martinussen R, Fletcher J, Cucchiaro G, Dalgalarrondo P, Lui M, Tannock R (2004) The Strengths and Difficulties Questionnaire overseas: Evaluations and applications of the SDQ beyond Europe. European Child and Adolescent Psychiatry 13:II/47-II/54 (Supplement 2).

www.sdqinfo.com

www.corc.uk.net

Zahner G, Pawelkiewicz W, DeFrancesco JJ, Adnopoz J (1992) Children's mental heath service needs and utilization patterns in an urban community: an epidemiological assessment. Journal of the American Academy of Child and Adolescent Psychiatry 31:951-060.

APPENDICES

Appendix 1. Tables for comparison of Finnish and international SDQ studies

Table 15. Internal consistency of the SDQ parent-reported scales (Cronbach's alpha) in different studies.

Studies	Koskelainen et al. 2000b	Smedje et al. 1999	Goodman 2001	Malmberg et al. 2003	Muris et al. 2003	Widenfelt et al. 2003	Worner et al. 2004	
Samples	community n = 703	community n = 900	community n = 9998	community n = 263	community n = 562	community n = 300	community n = 930	
Age range, years	7-16	6-10	5-15	5-15	9-15	8-16	6-16	
Alpha	alpha	alpha	alpha	alpha	alpha	alpha	alpha	average
SDQ								
Total difficulties	0.71	0.76	0.82	0.84	0.80	0.81	0.82	0.79
Emotional symptoms	0.69	0.61	0.67	0.71	0.70	0.68	0.66	0.67
Conduct problems	0.59	0.54	0.63	0.52	0.55	0.62	0.60	0.58
Hyperactivity -inattention	0.73	0.75	0.77	0.75	0.78	0.84	0.76	0.77
Peer problems	0.64	0.51	0.57	0.73	0.66	0.57	0.58	0.61
Prosocial behaviour	0.68	0.70	0.65	0.67	0.68	0.57	0.68	0.66

Table 16. The Internal consistency of the SDQ teacher-reported scales (Cronbach's alpha) in different studies.

Studies	Koskelainen et al. 2000b	Goodman 2001	Wiedenfelt et al. 2003	
Samples	community n = 376	community n = 7313	community n = 208	
Age range, years	7-12	5-15	8-12	
Alpha	alpha	alpha	alpha	average
SDQ				
Total difficulties	0.71	0.80	0.88	0.80
Emotional symptoms	0.79	0.65	0.76	0.73
Conduct problems	0.72	0.69	0.77	0.73
Hyperactivity – inattention	0.85	0.82	0.89	0.85
Peer problems	0.73	0.72	0.74	0.73
Prosocial behaviour	0.86	0.74	0.81	0.80

Table 17. Internal consistency of the SDQ self-reported scales (Cronbach's alpha) in different studies.

Studies	Koskelainen et al. 2000b	Goodman et al. 1998	Goodman 2001	Wiedenfelt et al. 2003	Muris et al. 2003	
Samples	community n = 528	community n = 83	community n = 3983	community n = 1288	community n = 562	
Age range, years	9-16	11-16	11-15	11-16	9-15	
Aplha	alpha	alpha	alpha	alpha	alpha	average
SDQ						
Total difficulties	0.71	0.82	0.80	0.70	0.78	0.76
Emotional symptoms	0.69	0.75	0.66	0.63	0.71	0.69
Conduct problems	0.57	0.72	0.60	0.47	0.45	0.56
Hyperactivity- inattention	0.66	0.69	0.67	0.66	0.72	0.68
Peer problems	0.63	0.61	0.41	0.39	0.54	0.52
Prosocial behaviour	0.69	0.65	0.66	0.60	0.62	0.64

Table 18. Inter-rater agreement between the parent- and the teacher-rated SDQ scales in different studies (Pearson's correlation coefficient).

Studies	Koskelainen et al. 2000b	Goodman 1997	Goodman et al. 1998	Goodman 2001	Wiedenfelt et al. 2003	
Samples	community n = 376	community n = 128	community n =83	community n = 7313	community n = 131	
Age range, years	7-12	4-16	11-16	5-15	8-12	
Correlation	correlation	correlation	correlation	correlation	correlation	average
SDQ						
Total difficulties	**0.44***	**0.62**	**0.43**	**0.46****	**0.52**	**0.49**
Emotional symptoms	**0.33**	**0.41**	**0.38**	0.27	**0.32**	**0.34**
Conduct problems	**0.30**	**0.65**	**0.28**	**0.37**	**0.36**	**0.39**
Hyperactivity-inattention	**0.45**	**0.54**	**0.36**	**0.48**	**0.54**	**0.47**
Peer problems	**0.39**	**0.59**	0.14 ns	**0.37**	**0.33**	**0.36**
Prosocial behaviour	**0.29**	**0.37**	0.16 ns	0.25	0.23	0.26

* All Finnish correlations were significant at level p < 0.01; **all correlations in Goodman 2001 study are significant at level p < 0.001; ns = not significant. The correlations, which are presented in bold in Table 19, were higher than the correlations found in the meta-analytic study by Achenbach et al. (1987).

Table 19. Inter-rater agreement between the parent- and the self-rated SDQ scales in different studies (Pearson's correlation coefficient).

Studies	Koskelainen et al. 2000b	Goodman 1998 community	Goodman 2001	Wiedenfelt et al. 2003	Klasen** et al. 2000	
Samples	community n = 528	community n = 83	community n = 3983	community n = 227	community n = 110	
Age range, years	9 -16	11-16	11-15	8-12	12-13	
Correlation	correlation	correlation	correlation	correlation	correlation	average
SDQ						
Total	**0.40***	**0.43**	**0.48**	**0.47**	**0.60**	**0.48**
Emot	**0.28**	**0.52**	**0.37**	**0.24**	**0.59**	**0.40**
Conduct	**0.28**	**0.36**	**0.44**	**0.41**	**0.36**	**0.37**
Hyper	**0.39**	**0.29**	**0.41**	**0.56**	**0.64**	**0.46**
Peer	**0.39**	**0.29**	**0.40**	**0.26**	**0.57**	**0.38**
Prosocial	**0.37**	**0.31**	**0.30**	0.14 ns.		**0.28**

*All Finnish correlations were significant at level p < 0.001; ns = not significant (p > 0.01); Klasen et al. (2000) used Spearman's rho coefficients. The correlations, which are presented in bold in Table 19, were higher than the correlations found in the meta-analytic study by Achenbach et al. (1987).

Table 20. Inter-rater agreement between the teacher- and the self-rated SDQ scales in different studies (Pearson's correlation coefficient).

Studies	Koskelainen et al. 2000b	Goodman et al. 1998	Goodman 2001	Wiedenfelt et al. 2003	
Samples	community n = 376	community n =83	community n =2767	community n =98	
Age range, years	9-16	11-16	11-16	8-12	
Correlation	correlation	correlation	correlation	correlation	average
SDQ					
Total difficulties	**0.38***	**0.38**	**0.33****	**0.29**	**0.35**
Emotional symptoms	**0.25**	**0.31**	**0.21**	0.12 ns.	**0.22**
Conduct problems	**0.30**	0.19 ns.	**0.30**	**0.39**	**0.30**
Hyperactivity-inattention	**0.34**	0.13 ns.	**0.32**	**0.44**	**0.31**
Peer problems	**0.38**	**0.24** ns.	**0.29**	0.16 ns.	**0.27**
Prosocial behaviour	**0.28**	0.19 ns.	**0.23**	0.19 ns.	**0.22**

*All Finnish correlations were significant at level p < 0.001; **all correlations in this study were significant at level p < 0.001; ns = not significant. The correlations, which are presented in bold in Table 19, were higher than the correlations found in the meta-analytic study by Achenbach et al. (1987).

Table 21. Correlations between parent-rated SDQ and CBCL (Pearson's correlation coefficient).

Studies	Koskelainen et al. 2000b	Goodman and Scott 1999	Wiedenfelt et al. 2003	Muris et al. 2003	Klasen** et al. 2000
Samples	community	psychiatric and dental	community	community	community
	n = 81	n = 132	n = 300	n = 562	n = 110
Age range, years	15-16	4 -7	8-16	9-15	12-13
Correlation	correlation	correlation	correlation	correlation	correlation
SDQ/ CBCL					
Total/total	0.75*	0.87*	0.74*	0.70*	0.78*
Conduct/Externalizing	0.70	0.84	0.72	0.60	0.60
Conduct/Delinquent	0.60		0.56		
Conduct/Aggressive	0.69		0.71		
Hyperactivity/Attention problems	0.67	0.71	0.78		0.79
Emotional/Internalizing	0.44	0.74	0.70	0.70	0.69
Emotional/Withdrawn	0.34		0.52		
Emotional/Somatic	0.40		0.45		
Emotional/Anxious-depressive	0.44		0.70		
Peer problems/Social problems	0.41	0.59	0.51		0.61

*all correlations were significant at p < 0.001; Klasen et al. (2000) used Spearman's rho coefficients.

Table 22. Correlations between the self-rated SDQ and the YSR (Pearson's correlation coefficient).

Studies	Koskelainen et al. 2000b	Wiedenfelt et al. 2003	Klasen** et al. 2000	Muris et al. 2003
Samples	community	community	community	normal
	n = 129	n = 300	n = 110	n = 562
Age range, years	15-16	8-16	12-13	9-15
Correlation	correlation	correlation	correlation	correlation
SDQ/ YSR				
Total/total	0.71*	0.70*	0.77*	0.74*
Conduct/Externalizing	0.68	0.62	0.59	0.56
Conduct/Delinquent	0.60	0.58		
Conduct/Aggressive	0.64	0.57		
Hyperactivity/Attention problems	0.59	0.66	0.78	
Emotional/Internalizing	0.68	0.64	0.73	0.74
Emotional/Withdrawn	0.43	0.48		
Emotional/Somatic	0.58	0.46		
Emotional/Anxious-depressive	0.68	0.65		
Peer problems/Social problems	0.51	0.41	0.58	

* all correlations were statistically significant (p < 0.001); Klasen et al. (2000) used Spearman's rho coefficients.

Table 23. The parent-, teacher- and self-rated SDQ total difficulties score means (sd) in different age and gender groups in community samples.

Studies	Koskelainen et al. 2000b	Koskelainen et al. 2000b	Wiedenfelt et al. 2003	British	British
Samples	community*	community	community**	community***	community
Age range, years	7-12	13-16	8-16	5-10	11-15
Correlation	correlation	correlation	correlation	correlation	correlation
Gender	boys/ girls	boys/ girls	boys/ girls	boys/ girls	boys/ girls
SDQ					
Parent-rated	6.7 (4.4)/ 6.2 (4.7)	5.3 (3.9)/ 5.3 (4.9)	7.8 (5.5)/ 5.6 (4.8)	9.3 (6.0)/ 7.9 (5.4)	8.8 (5.9)/ 7.6 (5.6)
Self-rated	9.6 (5.4)/ 8.3 (6.0)	7.5 (5.1)/ 7.0 (4.5)	9.7 (4.7)/ 10.1 (5.0)		10.5 (5.1)/ 10.0 (5.3)
Teacher-rated	7.8 (6.3)/ 4.8 (5.5)		9.1 (7.0)/ 5.8 (5.7)	8.0 (6.2)/ 5.6 (5.3)	7.6 (6.5)/ 5.0 (5.4)

* For the SDQ parent report n= 703; self-report n = 528, age range 9-16 years; for the teacher report n = 376, age range 7-12.
** For the SDQ parent report n = 300, age range 8-16; self-report n = 1353, age range 11-16; teacher report n = 208, age range 8-12 years. *** presented at www.sdqinfo.com, the SDQ parent report for 5-10-year-olds n = 5055, for 11-15-year-olds n = 4443; the teacher report for 5-10-year-olds n = 4801, for 11-15 year-olds n = 3407, for the self-report 11-15-year-olds n = 4228.

Table 24. The means of the SDQ self-report total difficulties scores and subscores in different studies.

Studies	Koskelainen et al. 2001b				Rønning et al. 2004				Van Roy et al. 2006		British*	
Samples	n = 1458				n = 2028				n = 9387		n = 2093	
Grade	7th	7th	9th	9th	7th	7th	9th	9th				
Age, years	13-14	13-14	15-17	15-17					13-16	13-16	11-15	11-15
Sex	boys	girls	boys	girls	boys	girls	boys	girls	boys	girls	boys	girls
SDQ												
total	10.5	11.6	11.0	11.8	9.6	8.8	9.9	10.1	10.8	11.1	10.5	10.0
hyper-activity	3.4	3.6	3.5	3.8	3.7	3.1	4.0	3.8	4.2	4.2	3.9	3.6
Emotional sympt.	2.2	3.4	2.4	3.6	2.0	2.6	1.9	3.1	2.1	3.2	2.6	3.0
Conduct problems	2.5	2.2	2.6	2.3	2.0	1.5	2.3	1.6	2.5	2.0	2.4	2.0
Peer problems	2.5	2.4	2.5	2.2	2.0	1.7	1.8	1.5	2.1	1.8	1.6	1.4
Prosocial behaviour	6.1	7.1	5.9	7.5	6.8	7.9	6.1	7.8	6.5	7.7	7.5	8.5

* British SDQ score distributions in www.sdqinfo.com.

Table 25. Cut-off points for normal and clinical range of the SDQ self-report scales in different studies.

Studies	Koskelainen et al. 2001b	Rønning et al. 2004	Rønning et al. 2004	Van Roy et al. 2006	Van Roy et al. 2006	British***
Percentiles	abnormal 90th percentile (%)*	borderline 80-90th percentile (%)	abnormal above 90th (%)	borderline 80-90 percentile (%)	abnormal above 90th (%)	abnormal above 90th (%)
Samples	community n = 1458	community n = 2028		community n = 9387		
Age range, years	13-17 years	11-16 years	11-16 years	13-16 years	13-16 years	11-15 years
SDQ						
Total difficulties	18 (9.4 %)	15 (8.9 %)	18 (9.0 %)	16-18 (8.8 %)	19 (10.4 %)	17 (9.2 %)
Emotional symptoms	5 (12.1 %)	5 (7.1 %)	6 (9.4 %)	5 (7.5 %)	6 (12.2 %)	5 (11.2 %)
Conduct problems	4 (12.4 %)	4 (7.8 %)	5 (7.0 %)	4 (8.5 %)	5 (11.7 %)	4 (10.6 %)
Hyperactivity	6 (8.2 %)	6 (8.6 %)	7 (10.0 %)	7 (7.7 %)	8 (7.8 %)	6 (11.5 %)
Peer problems	4 (9.8 %)	4 (7.0 %)	5 (8.8 %)	4 (7.1 %)	5 (10.1%)	3 (9.2 %)
Prosocial behaviour**	4 (10.8 %)	5 (11.5 %)	4 (8.3 %)	5 (11.4 %)	4 (9.7 %)	5 (8.6 %)

* percentage of all subjects scoring at or above 90th percentile; ** 0n the prosocial subscale, the cut-off point of 10th percentile was used and the percentages show those scoring at or below the 10th percentile *** British SDQ score distributions in www.sdqinfo.com.

Appendix 2.1. Receiver Operating Characteristics (ROC) analysis, sensitivity and specificity, and choosing an optimal cut-off point for a questionnaire

A questionnaire is usually used with a given cut-off point which separates the likely cases from the likely non-cases (normals), i.e. those individuals who score above or below the cut-off point. The result is a 2 x 2 contingency table, which can be summarized by two proportions: the percentage of "true" cases (by chosen criterion) scoring above the cut-off (true positive rate or sensitivity), and that of "true" normals scoring below that cut-off (true negative rate or specificity). As we move the cut-off point along the scale values, we obtain different 2 x 2 tables with varying values of true positive and true negative rates (Fombonne 1991).

The criterion-referenced validity of a questionnaire can be measured by its ability to correctly identify those children who are regarded as disordered (cases) by a criterion, e.g. child psychiatric interview (clinical diagnoses given) or a referral to mental health services (Verhulst and Koot 1992).

Relation between questionnaire scores and chosen criterion:

	Questionnaire scores		
	At or above the cut-off point	**Below the cut-off point**	**Total**
According to the criterion			
"True" case	a	b	a + b
"True" normal	c	d	c +d
Total	a + c	b + d	a + b + c + d

The sensitivity of a questionnaire, or a true positive rate, is the number of "true" cases who score at or above the questionnaire's cut-off point, divided by the total number of cases (a/ a + b).

The specificity of a questionnaire, or true negative rate, is the number of "true" normals who score below the cut-off point, divided by the total number of normals (d/ c+ d).

The false negative rate is the number of "true" cases who score below the cut-off point (b/ a + b).

The false positive rate is the number of "true" normals who score above the cut-off point (c/ c + d) (Verhulst and Koot 1992).

A Receiver Operating Characteristics (ROC) curve graphically displays the group of 2 x 2 contingency tables derived from all possible cut-off points of a scale. ROC analysis was first developed in signal detection applications as a means of separating the signal from noise. Then it was used later in medical applications and in psychiatric research. It is a quantitative method for assessing the overall performance of an instrument for a full range of its scores. The ROC curve is obtained by plotting on a graph the true positive rate on the vertical axis and the false positive rate on the horizontal axis. In the case of a perfect scale, no overlap would be observed between the distribution of scores among normals and cases. In this case, the corresponding ROC curve would be the upper-left part of the diagram. In contrast, two completely overlapping distributions would mean a useless scale. Equal proportions of normals and cases would be found above and below each score, thus leading to equal rates of true positives and false positives. In this case,

the ROC curve would be the diagonal of the diagram, also called the chance line or random ROC (Fombone 1991).

Several methods of accuracy have been proposed in ROC analysis. The most commonly used is the area under the curve (AUC). This area was the dimensions of a probability, and it varies between 0.5 (which corresponds to the AUC of the diagonal) and 1.0 (which is the AUC of a perfect instrument). On the diagram, the further upward and to the left the ROC curve deviates from the diagonal, the better is the performance of the instrument. The intuitive interpretation of the AUC is that it is the probability of correctly classifying a randomly selected pair of subjects, where one is "normal" and one is "case", however a case is defined. (Fombonne 1991; Kresanov et al. 1998; Verhulst and Koot 1992; Hanley and McNeil 1982).

Appendix 2.2. Selection of the optimal cut-off point

When using questionnaires in research and clinical practice the choice of a cut-off point for the scores affects the accuracy of the results. Goodman (1997) discusses the use and choice of an optimal cut-off score. The percentile scores may be used, e.g. if the researcher wants to know whether a sample of girls with high SDQ scores is comparable to age- and sex-matched controls. However, originally Goodman proposed the cut-off points of the SDQ scores to act as useful "warning signals" for the clinician that the adolescent may have a disorder (Goodman 1997; Goodman et al. 1998). Thus, cut-off points may act as markers for preventive or clinical action.

Goodman et al. (1998) presented normal, borderline and abnormal provisional bandings for the SDQ self-report. They found that with the cut-off point of the SDQ total difficulties, 31 % of the clinical sample, were in the abnormally high range as compared with 5 % of the community sample. In addition, of the clinical sample, 28 % were in the borderline range, as compared with 18 % of the community sample. The remaining 41 % of the clinical sample scored in the normal range, as compared with 77 % of the community sample. Among those young people with normal or borderline self-report scores, who had also been rated by both parent and teacher SDQs, 72 % (18/25) of clinic attenders had high informant-rated total scores as compared with just 9 % (5/53) of community subjects.

Contrary to the SDQ studies where the cut-off points are sought and adjusted according to different samples, for the Rutter questionnaires, the standard cut-off is often used. Using a single cut-off point for all studies has both advantages and disadvantages (Goodman 1997). The advantages are simplicity and equivalence across studies. The main disadvantage is that "caseness" will not have a comparable meaning in different studies simply because those studies have employed the same cut-off point. Comparability is particularly likely to be lost when high- and low-risk samples are contrasted. The best strategy for researchers may be to choose cut-offs according to a likely disorder rate in the sample being studied, and according to the relative importance for that study of false positives and false negatives.

Selection of the optimal cut-off point depends on the aims of the study, the prevalence rates of the studied conditions, the relative weight that the researcher gives to the two types of classification errors, false positive and false negative rates, as well as on the knowledge of the distributions of the scores in cases and in normals. For example, a more stringent cut-off is needed when selecting more severe cases and a less stringent one in choosing minor cases. (Kresanov et al. 1998; Fombonne 1991).

Screening in the first phase of a two-stage community survey should be achieved with sufficient efficiency. In cross-sectional surveys, one goal of the second stage is to estimate with good accuracy the sensitivity and specificity of the screening instrument. This allows for the derivation of an estimate of the "true" prevalence rate in the target population. However, second-stage interviews are more difficult and more costly to obtain. Thus, sampling the screen-positive group should result in an elevated probability of selecting a case for further investigation, in order to optimally use available interviewing resources. The cut-off point chosen for the first stage should yield a high enough positive predictive value (probability that a screen-positive is truly a case), and sometimes this will be achieved at the expense of a high false negative rate (Fombonne 1991).

Appendix 3. Studies of factor analysis for the SDQ parent report

The factor structure of the SDQ parent report has been studied using, e.g. Swedish (Smedje et al. 1999), German (Woerner et al. 2004), American (Dickey and Blumberg 2001) and Australian (Hawes and Dadds, 2004) versions of the SDQ. A factor analysis of the German version of the parent SDQ obtained from a community sample of 6-16-year-olds (n = 930) showed an exact replication of the original five scales (Woerner et al. 2000). However, the results of the Swedish and American studies are somewhat different.

In the Swedish study by Smedje et al. (1999), parental ratings of 900 children aged 6-10 years from the general population were gathered. They found that factor analysis generally confirmed the subscale structure for both sexes, although an exception was found for boys, when one emotional symptom scale item (often complains of headaches, stomach-aches or sickness) loaded weakly on the conduct problem and peer problem subscales, contrary to the findings for girls. In addition, for girls, one conduct problem item (often has temper tantrums and a hot temper) loaded most strongly on the emotional symptoms.

Dickey and Blumberg (2004) made a structural analysis of the parent SDQ. Parents or guardians of a national probability sample of 9574 children and adolescents aged 4-17 years filled in the questionnaire. Principal component analysis (PCA, Promax-rotation allowing the components to correlate) suggested that a five-component solution would best explain the total variance. The results of the PCA revealed that the hyperactivity-inattention, emotional symptoms and prosocial problems components were identified by items that were intended to represent these domains. However, the components representing conduct problems and peer problems were represented by fewer than the intended five items. Although the majority of conduct problem items were related to the conduct dimension, three of the five items were also related to other components (lies and temper items to both conduct and hyperactivity; disobedient related, instead of to conduct problems, to hyperactivity and prosocial behaviour). The peer problem scale was identified only by two items (best with adults and solitary). Other items were related to prosocial problems (lack of a good friend and unpopular) or to emotional problems (bullied). However, utilization of the eigen values greater than 1 rule (Promax rotation), revealed that although three factors represented the optimal factor solution, one item (steals) was not related to any extracted factor. This item was removed but the three-factor solution still remained optimal. Altogether, these three factors may be interpreted as prosocial behaviour, externalization problems and internalization problems.