Strengths and Difficulties Questionnaire (SDQ)

Evaluations and applications

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von

Andreas Becker
aus Göttingen

Göttingen, den 19.03.2007

D7

Referent: Prof. Dr. Marcus Hasselhorn

Korreferent: Prof. Dr. Eckart Rüther

Tag der mündlichen Prüfung: 02. Mai 2007

ABSTRACT

This study examines whether the German translation of the Strengths and Difficulties Questionnaire (SDQ) is a reliable and valid screening instrument and whether it is as effective a tool for clinical diagnostics and scientific applications as the CBCL / TRF and to evaluate the German self-reported Strengths and Difficulties Questionnaire (SDQ) in a clinical setting. We also investigated whether this additional information gathered directly from older children and adolescents improves the prediction of clinical status when external ratings from their parents and/or teachers are already available. We examined 543 children and adolescents (147 girls and 396 boys) with ages ranging from 5 to 17 years and correlated the results of the parent and teacher SDQ as well as the CBCL / TRF with clinical diagnoses. Furthermore, the adequacy of the scale structure of the SDQ was tested using confirmatory and exploratory factor analyses. Additionally, a sample from 10 European countries with 1,459 children with ADHD (aged 6–18 years) will be analyzed. It was demonstrated that the scales of the parent and teacher versions were sufficiently homogeneous (.72 - .83). Correlations between SDQ scales and corresponding CBCL / TRF scales showed a high degree of congruence, while an exact replication of the original SDQ scale structure could also be achieved. Parent and teacher versions of both questionnaires presented with good validity, not only with regard to the discrimination between child psychiatric patients and a representative community sample, but also in the identification of different categories of disorders within the clinical sample. The mean total difficulties and SDQ subscale scores of the ADORE sample clearly differed from UK normative data. Younger children were more impaired on different SDQ scales than older children, and girls were more emotionally affected than boys. Differences between countries were found for each SDQ scale, but the investiga! tor type had no significant effect. Correlation coefficients between S DQ scales and other scales

used in ADORE ranged from low (r<0.30) to high (r>0.50). The parent and teacher and self-rated version of the SDQ proved to be valid and helpful questionnaires for use in the framework of a multi-dimensional behavioural assessment, and appear to be well-suited for screening purposes, longitudinal monitoring of therapeutic effects, and scientific research purposes. In contrast to investigator type, different cultures had a significant effect on SDQ scores. Correlations with other scales used in the ADORE study underline both separate domains and meaningful associations.

ZUSAMMENFASSUNG

Diese Studie untersucht, inwieweit die deutsche Übersetzung des Strengths and Difficulties Questionnaire (SDQ) ein reliables und valides Untersuchungsinstrument ist. Dieses soll sowohl für den klinisch diagnostischen Bereich als auch für wissenschaftliche Fragestellungen im Vergleich zu der CBCL/TRF geprüft werden. Weiterhin soll ebenfalls die Selbstversion des SDQ hinsichtlich seiner Gütekriterien überprüft werden. Zusätzlich ist von Interesse, ob die Vorhersage bzgl. des klinischen Status sich verbessert, wenn ergänzend zu der Eltern und/oder Lehrerinformation die Selbstauskunft der Jugendlichen ebenfalls berücksichtigt wird. Wir untersuchten 543 Kinder und Jugendliche (147 Mädchen und 396 Jungen) im Alter von 5-17 Jahren und verglichen die Ergebnisse des Eltern und Lehrer SDQ mit der CBCL und TRF und den jeweiligen klinischen Diagnosen. Darüber hinaus wurde die Skalenstruktur mittels einer konfirmatorischen und explorativen Faktorenanalyse getestet. Außerdem wurde eine europäischen Stichprobe von 1459 Kindern, die eine ADHD Diagnose erhalten haben (6 -18 Jahren), mittels des SDQ untersucht. Es konnte gezeigt werden, dass die Eltern- und Lehrerskalen eine ausreichende Homogenität (.72-.83) aufwiesen. Korrelationen zwischen SDQ-Skalen und korrespondierenden CBCL/TRF Skalen zeigten eine hohe Übereinstimmung. Ebenso konnte eine exakte Replikation der intendierten SDQ-Skalenstruktur aufgezeigt werden. Die Eltern und Lehrerversion der beiden Fragebögen zeigten eine gute diskriminante Validität, nicht nur im Hinblick auf die Unterscheidung zwischen verschiedenen kinderpsychiatrischen Störungen in einer Klinikstichprobe, sondern ebenfalls innerhalb einer repräsentativen Feldstichprobe. Der Totalproblemwert der SDQ-Subskalen innerhalb der ADORE-Stichprobe unterschied sich klar von den Normwerten der englischen Stichprobe. Jüngere Kinder waren beeinträchtigter auf vielen SDQ-Skalen als ältere Kinder und Mädchen waren mehr emotional

beeinträchtigt als Jungen. Unterschiede zwischen den einzelnen Ländern wurde für jede SDQ-Skala gefunden, wobei die Art der Befragung auf diese Unterschiede keine Auswirkung hatte. Zusammenhänge zwischen den SDQ-Skalen und anderen Fragebogeninventaren innerhalb der ADORE-Stichprobe bewegten sich zwischen $r < .30$ und $r > .50$. Die Eltern-, Lehrer- und Selbstversion des SDQ wurde evaluiert um die Brauchbarkeit des Fragebogens innerhalb einer multidimensionalen Bewertung von kinderpsychiatrischen Auffälligkeiten zu überprüfen und zeigte, dass der SDQ ein valides Instrument für Screeningzwecke, für Längsschnittuntersuchungen, für therapeutische Effekte und für wissenschaftliche Fragestellungen ist. Im Unterschied zum Typ der Befragung konnte gezeigt werden, dass verschiedene Kulturen einen bedeutsamen Einfluß auf die Höhe der SDQ Skalenwerte hatten.

TABLE OF CONTENTS **Page**

1. BACKGROUND

The aim of this cumulative dissertation in the context of four original papers on the validation and application of the German Strengths and Difficulties Questionnaire (SDQ) is to provide a survey on the studies published to date on the psychometric quality of the SDQ. The following paper will describe the various application areas of the SDQ in different scopes of application and countries. Finally, aspects of culture comparative research using the SDQ will be discussed.

The main goal of this paper is to present published original papers on the SDQ. The first of these papers deals with the aspect of clinical validity in a German clinical sample in which the SDQ was used. This study examines whether the German translation of the Strengths and Difficulties Questionnaire is a reliable and valid screening instrument and whether it is as effective a tool for clinical diagnostics and scientific applications as the CBCL / TRF. The aim of the second study was to evaluate the German self-reported Strengths and Difficulties Questionnaire in a clinical setting. We also investigated whether this additional information gathered directly from older children and adolescents improves the prediction of clinical status when external ratings from their parents and/or teachers are already available. The third article deals with the question as to the clinical value of the parent SDQ for the child psychiatric practice. It is essential to establish whether a mental disorder requiring treatment is present and whether the subscore hyperactivity/attention deficit represents and aid for ADHS diagnostics. The fourth and last article concerns the question as to the psychometric properties in a large European sample. To examine the psychometric properties of the SDQ parent version and to determine the effects of age, gender, country and investigator type (paediatrician, child psychiatrist, other

physician) on the SDQ scores in the prospective, non-interventional ADORE study in children with ADHD.

2. INTRODUCTION

With the worldwide increasing awareness of mental health problems in children and its high impact on public health issues as well as the economic future of the countries, it is of utmost importance to have empirically tested inventories at hand to measure psychopathology in a standardized way. This would allow to optimize assessment, diagnostic grouping and treatment, either for coordinated worldwide multi-site research and/or regional development of efficient mental health services.

To reach this goal, the inventories in question must reflect a similar range of psychopathology in different societies but at the same time being sensitive enough to detect differences between cultures to tune the necessary regional mental health support for the well being of the local patients.

2.1 General Aspects

Psychopathology can be *defined* as the systematic study of abnormal behaviour, experiences and cognitions (Sims, 2003). In descriptive psychopathology, an attempt is made to describe and categorize the observed abnormal behaviour and the abnormal experiences and cognitions as reported by the patient or a proxy in the case of younger children. Descriptive psychopathology may also be equated with the term phenomenological psychopathology because it is reflecting the various phenomena of observable abnormal behaviour and internal experiences and cognitions of the patient. Both the observable behaviours and the internal experiences and cognitions are considered to be symptoms if they clearly reflect abnormal phenomena.

2.1.1 Nosological systems of classification

The various psychopathological symptoms are *grouped* into disorders with the latter being organized in nosological systems of classification. There is a long-standing history of setting up different schemes of classification. In the past, in psychiatry of all ages there has been an emphasis on local schools of thought leading to a large variety of local or national nosological concepts. However, this traditional understanding has come to a halt in the second half of the last century and gave way to the development of more universal and international schemes of classification ending up in the current two major international systems of ICD-10 and DSM-IV.

This tradition is rooted in the understanding that psychopathology is best *reflected* in different nosological categories. The categorical approach has the advantage that it is most suitable for communication both with layperson and among scientists and reflects a tendency of the human mind to organize complex information. However, it should be clear that neither medicine in general nor psychiatry specifically fully adheres to the categorical model. For instance, hypertension as a categorical disease is based on a quantitative elation of blood pressure which is distributed dimensionally in the population. Similarity, despite categorical definition Attention-Deficit Hyperactivity Disorder (ADHD) is considered as a dimensional trait with different severity of expression in the population in many contemporary genetic research studies.

Various *instruments* for the assessment of psychopathological symptoms and clinical disorders have been delineated from current schemes of *classification*. Particularly for research, attempts have been made to standardize the interviewing process in order to arrive at reliable and valid assessment procedures that reflect both

psychopathological symptoms and diagnostic criteria of the various disorders. Examples include highly structured interviews like the Diagnostic Interview Schedule for Children (DISC) (Fisher, 1993) or semistructured interviews like the Parental Accounts of Symptoms (PACS) (Taylor, 1986) and the Parent Interview for Child Symptoms (PICS) (Ickowicz, 2006).

2.1.2 Dimensional approach of ordering symptoms

An alternative view to classification is represented by the *dimensional approach* of ordering symptoms based on multivariate statistical approaches to classification. Since the early studies in the middle of the last century, statistical techniques have been repeatedly applied in order to isolate interrelated patterns of behaviour in children and adolescents. In the middle of the eighties of the last century, there had been more than sixty studies spanning almost 40 years in which this approach had been used (Quay, 1986). Four major dimensions had been established most clearly, namely, undersozialized aggressive conduct disorder, socialized aggressive conduct disorder, attention deficit disorder, and anxiety-withdrawn-dysphoria. The empirical foundation of two additional dimensions, namely, schizoid-unresponsive and social ineptness was less firm and the dimension of psychotic disorder was difficult to identify in sufficient quantity for multivariate statistics due to the rare manifestation of psychotic disorders in the childhood population. There is a direct line from this historical development of dimensional classification to the contemporary use of questionnaires for the assessment of emotional and behavioural abnormalities both in clinical practice and research.

2.2 Questionnaires

Various *methods* are used to obtain information for diagnostic assessment and classification of individual psychopathology and behaviour. Besides the clinical interview, parent- and teacher-completed checklists, child- or adolescent-completed self-reports, self-monitoring of behaviour, information from third parties, and direct observation can be used. It is good clinical practise to incorporate multiple methods across multiple informants in order to arrive at a comprehensive picture of the various facets of individual psychopathology. In this process of assessment, questionnaires are an efficient tool for gathering standardized information.

2.2.1 Different purposes of questionnaires

Questionnaires or rating scales serve different purposes, including screening for emotional and behavioural problems, assisting in the diagnostic process, and monitoring changes in behaviour across time or as an effect of intervention. Besides the issue of classification of abnormal behaviour, research has made use of questionnaires when it came to the identification of at-risk individuals in large epidemiological surveys. In this process, many questionnaires or rating scales have been developed in the last decades with some measuring general psychopathology like the Child Behaviour Checklist (CBCL) (Achenbach, 1991) or the Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997) and others aiming at the assessment of specific disorders and behaviours. Numerous scales are available e.g for the assessment of ADHD, anxiety disorders, autism, depression, or obsessive-compulsive disorders. These questionnaires for specific disorders will not be considered in the present study.

Ratings for child and adolescent psychopathology have to be age-related in order to consider the *developmental effect* on symptoms and behaviour. As a consequence, the composition of items and even the dimensional structure of the instrument may change with age. Furthermore, age-adjusted norms may be necessary in order to allow valid comparisons of a given child with the standardization sample. The same may also apply to *gender* because there are gender differences with regard to the manifestation of certain behaviours in children, e.g. aggressive or emotional symptoms. Scoring procedures and interpretations should be easy to understand and preferably guided by computer programs including graphic displays.

Furthermore, given the informant and situational specificity of behaviour it is of major importance to get information from more than one *source.* Rating scales for the assessment of general child and adolescent psychopathology like the CBCL or the SDQ fulfil these expectations by having parallel versions for parents and teachers and providing also parallel self-reports for older children and adolescents. Additional features of clinical utility include the availability of *translations* given the fact that a large proportion of patients are from different ethnic backgrounds than the indigenous clinical population in a certain country.

2.2.2 Psychometric quality of questionnaires

The *psychometric quality* of questionnaires and rating scales is a basic requirement for their application in clinical practice. Thus, proven reliability in terms of test-retest reliability over two different occasions and interrater reliability in terms of a satisfactory level of agreement between two informants is mandatory for a sound instrument. The latter should be taken from informants who actually have the same window of information and observation because situational specificity of behaviour

may jeopardize reliability. Furthermore, the internal consistency should clearly show that the items of a given scale are sufficiently interrelated.

The other major psychometric feature of validity may be more difficult to establish than reliability. Various terms and methods for validity are in use. Content validity refers to the fact that the items really reflect the behaviour that they are aiming to tap. Construct validity reflects the underlying concept of the scale by demonstrating that scores on a given scale are related to external criteria, such as course or response to treatment. Concurrent validity may be tested by calculating the correlation between scales with similar constructs. Criterion-related or predictive validity addresses the relationship of a scale to another measure, such as the diagnosis based on a clinical interview. Several studies have assessed this issue by use of the CBCL scales in the prediction of various clinical disorders. Sensitivity to change is another important validity aspect that helps in the implementation of clinical rating scales when it comes to the evaluation of therapeutic interventions. Clinicians need to be aware of the psychometric features of questionnaires in order to benefit from these tools in their practice.

2.2.3 Advantages and disadvantages of questionnaires

In summary, clinical questionnaires may be characterized by various advantages and disadvantages (Conners, 1998). First, there are a number of impressive *advantages* for clinical practice. Ratings can draw upon substantial previous experience with a child or adolescent over extended time intervals and diverse situations. Even rare and infrequent behaviours can be gathered that may be missed by interviewing. Questionnaires are cheap and extremely efficient in the time needed to collect the information. They need not be administered by trained professional staff. If normative

data are available, normalization of behaviour can be studied in a given child, e.g. as a consequence of intervention. Ratings of children's behaviour by parents, teachers and significant others have substantial ecological importance regardless of accuracy or reliability. Finally, ratings permit the quantification of qualitative aspects of behaviour in children that is not readily gathered by other means of assessment.

However, questionnaires have also various *disadvantages*. There may be various systematic rating errors like leniency or severity errors in the rating of a behaviour, halo effects with unfair positive or negative slant to all items upon the judgement of particular behaviours, logical errors, contrast errors depending upon the comparison with whom the subject is compared, and recency errors with a rating of the most recent episode of behaviour. Other disadvantages include the limitation to the informant's perspective and the limitation to the content of the scale. Furthermore, the informant's responses and subjective experiences may not be explored, direct behaviour observation is missing, and misunderstandings may not be clarified. Finally, rating scales are not good enough for definite clinical case identification because even with high sensitivity and specificity the number of false-positives will outnumber the true cases. Thus, good clinical practice must rest on a combination of assessment instruments including questionnaires and direct examination by use of interviewing.

Hence, this work focuses on the transcultural aspects of a worldwide accepted broad-band psychopathological questionnaire (SDQ) describing and setting into perspective their usefulness in a rapidly changing/exchanging globalizing world.

3. METHODS

3.1 The Strengths and Difficulties Questionnaire (SDQ)

The Strengths and Difficulties Questionnaire (SDQ) has within a decade become one of the most frequently used assessment tools in child and adolescent mental health research [(Goodman, 1997); (Rothenberger, 2004)]. The SDQ has been translated into more than 60 languages, and studies with the SDQ have been published from all continents. Publications include psychometric evaluation in different languages and cultures, epidemiological surveys, and assessment of at-risk groups of children and adolescents. As a short instrument including positive descriptions of the child, it is rapid to administer and well accepted even in non-clinical populations.

The SDQ evolved from a modified version of the Rutter Questionnaire which had been used extensively in early population surveys (Goodman, 1997). Some items were modified in order to form five subscales and include positive as well as negative descriptions of behaviour. It has 25 items, five each for the subscales Emotional symptoms, Conduct problems, Hyperactivity, Peer problems and Prosocial behaviour. The sum of the first four makes up the Total difficulties scale. The three response categories are 0 = Not true, 1 = Somewhat true and 2 = Certainly true. An Impact supplement is frequently added to the symptom rating, asking about distress to the child, burden to others, and interference with home life, friendships, classroom learning and leisure activities. There are versions of the SDQ for parents, teachers and self-report (the latter for age 11 and higher). It covers the age range 5-17 years, but a separate parent version exists for 3-4 year-olds. There are also versions for repeated assessment following treatment in the clinic (follow-up). Questionnaire, scoring instructions, an additionally computerised algorithms for predicting psychiatric disorder by bringing together information on symptoms and impact from SDQs

completed by multiple informants and scoring software are available without cost for non-profit use from the web-site www.sdqinfo.com.

3.2 Psychometric properties of the SDQ

In the course of the past few years, SDQs have been completed for more than 180,000 children and adolescents in population-based studies as well as in clinical samples. During this time, use of the SDQ has lead to the accumulation of a steadily growing body of empirical evidence in different parts of the world, reflecting a strong practical and scientific interest in this instrument. Therefore, this overview compiles past and current evaluations and applications of the SDQ, not only reporting a wider selection of European reports but also reviewing the current status of the SDQ overseas.

3.2.1 Reliability of the SDQ

The psychometric qualities of the SDQ have been assessed in various studies in different countries. The first of these studies (Goodman, 2001), evaluated the psychometric properties of the original version of the SDQ in a total of 10.438 British children aged 5 to 15 years. The *internal consistency* coefficients (Cronbach's α) for the parent-rated SDQ subscales and the total problem score were generally satisfactory (mean 0.73), particularly for the total difficulties and total impact scores (all 0.80 or higher). Cronbach's alpha coefficients for the teacher-rated version were fairly high for all scales. The lowest value was found for the subscale measuring peer problems (0.70) and the highest alpha coefficients were found for hyperactivity/ inattention (0.88) and prosocial behaviour (0.84) subscales. Thus, reliability of the parent-rated and teacher-rated version of the SDQ in this sample was very

satisfactory. However, the internal consistency of the self-report peer problems scale was only moderate (0.41 - 0.67).

In the course of the last 10 years several studies have shown that the SDQ scales provide a satisfactory to good internal consistency for different cultures. In one of the first studies in a Swedish non-clinical sample (Smedje, 1999) the Cronbach's alpha coefficients for reliability in the scales demonstrated a moderate to good consistency. A good consistency of the SDQ scale was found for these children who were rated by their parents. Similar results were reported in a Dutch study (Muris, 2003), in which healthy children and adolescents were surveyed. Again it was shown that the internal consistency for the various SDQ scales were generally satisfactory for the parent version (mean=0.70) and for the teacher version (mean=0.64). Only the consistency for the self report conduct problems (0.45) and peer problems (0.54) was notably low. A further investigation in a community sample from Australia (Hawes, 2004) showed a moderate to strong internal reliability across all SDQ scales in a parent-rated survey. The results of the German standardization of the SDQ (Woerner, 2002) showed that homogeneity of the SDQ scales was satisfactory to good. The Cronbach's alpha value was 0.82 for the entire scale, and the values for the individual subscales were 0.58 - 0.76. The internal consistencies obtained for adult informant-rated SDQ scales in this clinical sample were rated again as good. None of the internal consistency coefficients was lower than .70 (.72 - .81 for parent subscales; .75 - .83 for teacher subscales). For the total difficulties score based on 20 items, parent- and teacher-rated instruments yielded identical coefficients (0.83). Thus, both parent and teacher versions can be considered to be sufficiently reliable (Becker, 2004b). Recently, evidence of the good internal consistency of the SDQ was also found in a prospective/non-interventional study in 10 European countries in which 1459 children with the diagnosis of ADHD participated (Becker, 2006b). The

Cronbach's alpha coefficients were quite high in the evaluated sample. This was fairly consistent for all countries. Results on the internal consistency demonstrate the homogeneous scale structure, with reliabilities for the parent total difficulties score ranging between 0.82 (Goodman, 1998) and 0.71 (Koskelainen, 2000) and 0.76 (Muris, 2004) for the self version in several studies of different societies.

3.2.2 Test-Retest Stability of the SDQ

The *stability* of the SDQ scales were first examined by Goodmann and Scott (Goodman, 1999). It could be shown, that in a sample of British children very satisfactory results were found for intraclass correlation (ICC) in a time period of 3 to 4 weeks. For the total problem score a ICC of 0.85 was found. The ICC for the SDQ subscales was between 0.74 and 0.83 and a good all round test-retest reliability was observed. The shortest time period (2 weeks) for the evaluation of retest stability was chosen in the pilot study of Smedje (Smedje, 1999). In this investigation, it was found that the parent rating of problem behavior in their children was very stable (total score=0.96). The study by Muris (Muris, 2003) in the Netherlands also showed that test-retest stability of the SDQ over a 2-month interval was satisfactory (parent: 0.76 - 0.91; self: 0.59 - 0.88).; with the exception of the prosocial behaviour subscale of the self-report SDQ (ICC=0.59), all intraclass correlation coefficients were well in the 0.70 range or higher. A study by Goodman (Goodman, 2001) demonstrated a mean retest stability of 0.62 for the parent-rated SDQ despite a very long interval of 4 to 6 months. Teacher ratings were most stable (mean correlation=0.73) and youth ratings least stable (mean =0.51). In general, stability was greatest for the total difficulties and hyperactive-inattention scores.

3.2.3 Factor structure and factorial validity of the SDQ

In a nationwide epidemiological investigation of 5 to16 year–old children in the UK (Goodman, 2001) the postulated *factor structure* was studied in 10,438 children and a five-factor solution could be found by using a exploratory factor analysis (EFA) for the parent, teacher as well as self version. For all three rating categories, all 25 items loaded on the predicted factors, with a few items also loaded on additional factors. The factor structure of the German parent SDQ (Woerner, 2002) was also examined using an enlarged data set. This combined database (Becker, 2004b) was created by pooling the German standardization sample and the entire clinical validation sample with available parent SDQ data. Thus, a total of 1686 parent-rated SDQs from a mixed community and clinical sample were included. A high correspondence between the pattern of rotated loadings and the original SDQ scales was found. According to this, factor analysis of the German SDQ in a clinical sample further confirmed the original scale structure of the instrument. Additional analyses of the parent SDQ by means of confirmatory factor analysis showed a good model fit. The 25 parent-rated SDQ items were subjected to a confirmatory factor analysis (CFA), which demonstrated a good model fit (AGFI = .85) of the original 5-factor model and a sufficiently low error term (RMR = .07). The study from Sweden (Smedje, 1999) used parental ratings pooled from two samples. The primary result of the psychometric study was to provide empirical support for the postulated factorial structure of the SDQ. The same result was obtained in a 2003 Dutch study (Muris, 2003) for the parent and self version of the SDQ. Analysis of these 5-factors revealed that all items loaded convincingly on the intended factors of hyperactivity-inattention, emotional symptoms, peer problems, conduct problems, and prosocial behaviour. Only one item had a substantial secondary loading: the prosocial behaviour item "considerate" loaded on the conduct problems factor. A similar pattern emerged by

factor analysis of the self-report SDQ. However, the items *lies* and *tempers* that were supposed to represent conduct problems substantially loaded on the peer problems factor. A further conduct problems item "obedient" clearly loaded (negatively) on the hyperactivity/inattention factor. Finally, the item "good friend" that was thought to load negatively on the peer problems factor, loaded convincingly on the conduct problems factor. An Australian study (Hawes, 2004) on the psychometric properties of the SDQ in a community sample demonstrated that the five factor solution was consistent with the original subscales. For both genders, most items loaded moderately to strongly onto their predicted factors. Data from Yemen (Almaqrami, 2004) in the context of a validation study of the self version of the SDQ also confirmed the original five factor solution.

Although studies from many different countries demonstrate that the results of the scale structure significantly support the intended five factor model, it must be noted that some studies failed to replicate the predictive factors. In an Arabic study (Thabet, 2000) the authors concluded that while the original factors were somewhat evident in the sample, the individual subscale appeared to be more heterogeneous or multifactorial than observed in other populations. The study reported by Dickey and Blumberg (Dickey, 2004) in a large representative U.S. sample did not entirely confirm the predictive five-component structure. Some items intended to assess conduct problems were more closely related to hyperactivity, and some items intended to assess peer problems were more strongly correlated with emotional or prosocial problems. Factor analyses revealed a stable three-factor model consisting of externalization problems, internalization problems, and a positive construal factor. In another study which further pursued the psychometric properties of the self version of the SDQ (Muris, 2004) indicated that peer problem and conduct problem items did not load on separate factors. On the basis of these results, a four-factor solution

(emotional symptoms, hyperactivity-inattention, prosocial behaviour and a mixed factor of peer and conduct problems) was examined. Results for this solution were more satisfactory.

3.2.4 Content and criterion-related validity of the SDQ

Concerning the validity of the SDQ various aspects need to be considered. To assure that *content validity* is given for the SDQ, the items of the SDQ originally selected following factor analyses of more extensive screening inventories and are based on nosological concepts. In accordance with the classification systems of DSM-IV (APA, 1994) and ICD-10 (WHO, 1992), the items of the hyperactivity/inattention scale thus address overactivity (2 items), attention problems (2 items), and impulsivity (1 item). In contrast to similar questionnaires, special attention was given to a well-balanced inclusion of both positive and negative behavioural aspects, hence the raw scores of several items need to be inverted before they are summed up to scale values.

Another important psychometric characteristic is the *criterion-related or predictive validity,* the ability to differentiate between cases and non-cases. This question has already been dealt with in several studies. In a large UK data set (Goodman, 2001) it could be shown how strongly the various scales were associated with the presence or absence of psychiatric disorders. High SDQ values were connected with a considerably higher risk of a relevant DSM-IV (APA, 1994) diagnosis. Children and adolescents whose parent or teacher SDQ total difficulties scores placed them in the top 10% of the population ("abnormal range") were 15 times more likely to have a DSM-IV diagnosis; for self report, the corresponding risk was 6 times higher. Similar findings were reported in a Brasilian study (Fleitlich, 2001) in which SDQ information was gathered from a community sample, including ratings by their parents and

teachers. The percentage of positive and negative SDQ cases were then determined with respect to the independent clinical classification obtained with the DAWBA (Goodman, 2000a) results. It could be shown that there were statistically significant differences between the mean total scores (Woerner, 2004b). This difference was around one standard deviation for both adult informant-rated versions of the SDQ, but was much smaller for the self-report version. After adjusting for gender effects, scores of the clinical sample were found to be significantly higher than those of the combined community-based sample.

The report by Mullick & Goodman (Mullick, 2001) on Bangladeshi children demonstrated the good ability of different SDQ scales and informants to distinguish between community and clinic subjects. In the parent version it was observed that the scales for conduct problems (AUC=0.93), hyperactivity (AUC=0.92) and emotional symptoms (AUC=0.78) could provide a useful prediction. In contrast, the scales for total difficulties (AUC=0.64), peer problems (AUC=0.49) and prosocial behaviour (AUC=0.67) could not effectively differentiate between the clinical and community samples. The same was seen for the scale values of the parent SDQ in Yemen (Almaqrami, 2004) which was able to discriminate a sample of school children from a child psychiatric sample. The AUC-score of the total scores and subscale scores ranged from 0.77 to 0.89. The chance-corrected agreements between the clinical diagnosis and SDQ subscales prediction were significant. In a Pakistani study as well (Samad, 2005) it was found that the SDQ was able to discriminate between a group of psychiatric patients and a normative comparison group. In a further study reported by Hawes and Dadds (Hawes, 2004) of an Australian sample showed for each SDQ scale significant differences in prevalence between the high and low risk groups, indicating that higher scores were associated with a greater probability of being assigned a DSM-IV diagnosis. Higher values on

the SDQ scale "conduct problems" were significantly (OR=30.5) more often correlated with a diagnosis of "conduct disorder" and higher values on the SDQ scale "hyperactivity" were significantly (OR=17.9) more often correlated with a diagnosis of "any ADHD" were as assessed in a semi-structured telephone interview, than in children with lower SDQ values.

3.2.5 Relevance for the clinician

After having demonstrated the psychometric value of the SDQ, the question of *relevance for the clinician* is whether a patient has a mental disorder which requires treatment. This question was first examined by Goodman (Goodman, 2000b). SDQ predictions and independent psychiatric diagnoses were compared in a community sample. The algorithm makes separate predictions for three groups of disorders, namely conduct-oppositional disorders, hyperactivity-inattention disorders, and anxiety-depressive disorders. Each is predicted to be unlikely, possible or probable. Predictions of these three groups of disorders are combined to generate an overall prediction about the presence or absence of any psychiatric disorder. It was demonstrated that *multi-informant* (parents, teachers, older children) SDQs identified individuals with a psychiatric diagnosis with a specificity of 94.6% and a sensitivity of 63.3%. The questionnaires identified over 70% of individuals with conduct, hyperactivity, depressive and some anxiety disorders, but fewer than 50% of individuals with specific phobias, separation anxiety and eating disorders. Identical results were found in a 2004 study (Goodman, 2004) in which SDQ predictions and independent psychiatric diagnoses were compared in a community sample from a nationwide English survey. The sensitivity of the SDQ prediction of 'any diagnosis probable' depended on which diagnosis the child had, ranging from around 80% for anxiety and depressive disorders, to around 90% for conduct disorder and broadly-

defined attention-deficit/hyperactivity disorder, to near 100% for narrowly-defined hyperkinesis and less common disorders. Further one, the SDQ prediction works best when SDQs have been completed by both carers and teachers. The findings of this study suggest that screening with the SDQ could improve the detection and treatment of behavioural, emotional, and concentration problems among looked-after children.

One of the advantages of the SDQ is its impact supplement providing an estimate of burden, which is a part of the diagnostic criteria in child and adolescent psychiatry. It is probable that a combination of symptom and impact scores will be the best indicator of caseness also in other countries, as was the case for the British version of the SDQ (Goodman, 2000b) and the use of the extended version of the SDQ is recommended in future studies.

Moreover, the German version of the parent SDQ was tested (Becker, 2006a) in a clinical sample as to how well the SDQ and the subscale "hyperactivity" can contribute to diagnostic certainty in ADHD (Klasen, 2003); (Banaschewski, 2004). The most important question for the clinician is how many of the children who are rated by the parent SDQ as being borderline abnormal or abnormal are actually so, and how many of these children are not diagnosed as being abnormal (negative predictive value). The number of children diagnosed by a physician as having ADHD who at the same time were diagnosed as having a borderline abnormal or abnormal scale value on the subscale "hyperactivity" (sensitivity) or how many children without the diagnosis of ADHD have an inconspicuous scale value (specificity) were also investigated. For the entire group an "inconspicuous value" on the SDQ scale for "hyperactivity" was found in 85% (negative predictive value) of the cases without any concurrent clinical diagnosis of ADHD (correct negative classification). However, only

57% of the children who showed abnormal values on the scale for "hyperactivity" was diagnosed as having ADHD (positive predictive value).

Although the studies reviewed here represent only a selection among many standardization, validation and application studies concerning *SDQ* which have already been published, they provide a good basis for empirical evidence of these three questionnaires for epidemiological and clinical purposes in a great variety of societies.

3.3 Psychometric dialogue of CBCL/SDQ

In a further step the correlation between a semi-structured-interview (PACS) (Taylor, 1986) and these SDQ/CBCL scales were investigated. The interview-based ratings correlated significantly higher with the SDQ than with the CBCL scores, only for inattention-hyperactivity.

Further both CBCL and SDQ scores may predict the presence of a child psychiatric diagnosis (ICD-10, WHO 1992) as well as detect specific diagnostic subgroups in a clinical sample to the same extent (Becker, 2004b). Using the total difficulties score (SDQ) or total problems score (CBCL) to predict presence of any behavioural disorder, it could be shown that the SDQ as well as the CBCL were equally able to differentiate between patients with and without any clinical diagnosis on axis one. Moreover, the examined subscales of the CBCL and SDQ could also effectively detect more specific clinical subgroups, again revealing only minor differences between the two inventories. However, the hyperactivity-inattention scale of both the parent SDQ and the teacher SDQ allowed a more accurate prediction of the clinical subgroup with hyperactivity / attention-deficit disorders than the comparable attention problems scale of the CBCL and TRF. The composite internalizing problems score of

the parent CBCL was superior to the corresponding scale of the SDQ in detecting patients with emotional disorders (Becker, 2004b). Comparisons of the two adult informants' (parent, teacher) judgements revealed approximately equal predictive values for CBCL and SDQ scales.

In order to establish and compare the *criterion validity* of the Dutch self-rated version of the SDQ and the YSR with respect to detecting behaviour problems, logistic regression analyses were carried out using the respective scale score as a predictor of caseness (Muris, 2004). It could be shown that both the self-reported SDQ scores (OR=1.18 - 1.57) and the YSR (OR=1.11 - 1.28) discriminated well between the children who displayed behaviour problems at school and children who did not. The good *convergent validity* was also confirmed on the basis of the high correlation with the YSR scale values (*rs* between 0.14 and 0.61, p<.005).

3.3.1 Convergent validity of the SDQ

Also, good *convergent validity* (Becker, 2004b) could be seen between SDQ scales and their respective CBCL / TRF counterparts, including both total problem scores as well as the specific subscales. In particular, the conduct problems subscale of the SDQ showed a very strong concurrence (.82 for parent ratings and .86 for teacher reports) with the CBCL composite score for externalizing symptoms. Likewise, the 5 SDQ subscale assessing emotional symptoms was also highly correlated (.77 and .80 for parent and teacher-rated versions) with the global CBCL and TRF scores for internalizing symptoms. As expected, negative correlations were found between prosocial behaviour in the SDQ and the social problem scale of the CBCL / TRF. With respect to hyperactivity and attention problems, teacher judgements

demonstrated higher agreement between the two instruments compared than parent ratings did.

Moreover, the examined subscales of the SDQ and CBCL / TRF could also effectively detect more specific clinical subgroups, again revealing only minor differences between the two inventories. Analyses of *cross-informant* (Becker, 2004a) agreement yielded a similar correspondence between the self-rated SDQ and the parent version as had been found for the YSR.

3.4 Cross-cultural/globalizing aspects of the SDQ

The psychometric characteristics of SDQ have been tested in many countries and for all continents, in many clinical situations as well as in schools. While most studies have been carried out in the framework of epidemiological investigations, some studies also dealt with questions on the various aspects of validity or cultural comparison in the context of child psychiatric issues. The majority of studies have supported the five-factor structure and other psychometric properties of the SDQ in different populations. For these a substantive body of research exists on the psychometric properties of the Strengths and Difficulties Questionnaire in different cultures. These studies support the validity and reliability of its versions for the parent, teacher and self-reporting purpose, despite some variation in cut-off scores (Vostanis, 2006).

3.4.1 The use of the SDQ in Africa

According to estimates of the World Health Organisation (WHO) 10-20% of all children worldwide suffer from mental or behavioural problems. A study from Africa (Kinshasa), in a country in which little is known about the mental health of children, an investigation with a French version (teacher reported form) of the SDQ was

carried out in 1187 children (Kashala, 2005). Possible correlations between mental health problems und school performance, demographic factors, illness and nutrition were investigated. The results supported a five-factor structure and the research showed satisfactory internal reliability across all subscales (0.66-0.81), except for peer problems (0.35). The 90th percentile cut-off scores were somewhat different (higher) than the British published scores and varied according to gender. It is well known that the cultural/familial background of the person carrying out the rating has a decisive influence on the rating for abnormal behaviour. On the other hand, the higher SDQ cut-off scores, compared to the British scores, may indicate a higher frequency of mental problems among school children in Kinshasa. But it was found that the overall prevalence of problems among children in high-income countries such as Great Britain (Meltzer, 2000) and Sweden (Gillberg, 1998) varied between 10 und 20 percent and are similar to those seen in low-income countries such as India (Malhotra, 2002), Ethiopia (Ashenafi, 2001) and South Africa (Liang, 2002). Beware that the studies cited here use different approaches including different screening and diagnostic instrument, which probably contribute to the variation e.g. the high prevalence for Sweden. The advantage of comparing studies using similar instruments should be pointed out.

3.4.2 The use of the SDQ in the United States

A large study in the USA came to similar conclusions. The English version of the SDQ, after undergoing "Americanization" as authorized by Robert Goodman, found good acceptance among parents. The internal consistency of the SDQ scales corresponded with the English version (Goodman, 2001). Normative scoring bands were similar, though not identical, to the British bands. One of the methodological features of this study (Bourdon, 2005) was the use of three different scoring methods

for identifying children at high risk of disorder (i.e. high symptom scorers, parental perceptions of definite or severe difficulties and high symptoms plus impairment), with each scoring method being strongly associated with service utilization.

In the high total difficulties group, 45% of the children had already undergone professional child psychiatric treatment. Of the parent defined high difficulties group 59% of parents reported a service contact. In the high scale score plus impairment group, 56% had received counselling services. For cross-cultural comparison, the parent-defined high difficulties method was found to be advantageous. However, it should be considered, that there is often a higher correspondence between parent ratings and referral because parents are crucial for referral. Nevertheless, this does not necessarily mean that parent ratings are closer to clinical status after assessment, and that their ratings are more reliable in order to detect true cases. It allowed for parents of various cultural groups to evaluate their children's behaviour on the basis of their own concept of abnormality. Therefore, the American variation of the SDQ has been found to be an effective screening instrument to discriminate children who have undergone psychiatric treatment from those who have not.

3.4.3 Cross-cultural prevalence (USA/GB)

Another study (Mojtabai, 2006) compared the cross-cultural prevalence of emotional and behavioural problems among American and British children and adolescents. The samples for each country were taken in the framework of two large concurrent national surveys [(Green, 2005); (Simpson, 2005)]. Mojtabai compares the prevalence of serious emotional and behavioural problems, defined by high SDQ scale scores plus distress or impairment in role functioning in the two countries. First, it was observed that the SDQ possessed the same good psychometric qualities for

both countries. Furthermore, it was found that more British than American children met the criteria for emotional and conduct problems, but not hyperactivity/inattention. The prevalence was higher for all problems in young British boys and for emotional problems in older British girls. In spite of adjustment for minority status of the child and gender of the informant, the differences in the prevalence of emotional and behavioural problems persisted. These results corresponded with the findings of Bourdon (Bourdon, 2005) who found lower SDQ scores in the 2001 U.S. NHIS survey sample compared with those reported in the 1999 U.K. survey. The authors conclude that: "The modest differences between the American and British data may indicate that American parents judge their children somewhat more positively than do British parents. This may indicate that differing cultural perceptions and values exist for child behaviour problems. On the other hand, another study (McCarty, 1999) shows that there is no indication that the cultural background of parents influences their rating of problem behaviour in their children. Therefore, it is possible that the differences found actually reflect differences in disorder prevalence. Finally, no differences were found between countries in prevalence of contact with general medical providers.

3.4.4 The use of the SDQ in Scandinavian countries

The Strengths and Difficulties Questionnaire (SDQ) has been used also in the various Scandinavian countries (Sweden, Finland, Norway, Denmark & Iceland) and has also been utilized for the last 10 years in population-based studies as well as in clinical samples. The largest studies have been performed in Norway and Denmark, and in these countries the diagnostic interview DAWBA has also been used in conjunction with the SDQ. The descriptive statistics suggest (Obel, 2004) that the distributions of SDQ scores are very similar in all the Scandinavian countries. The

validity of the SDQ as a screening instrument is currently being tested in many countries and studies using psychiatric interviews such as the DAWBA (Goodman, 2000a) and K-SADS (Ambrosini, 2000). Thus, measures of sensitivity and specificity will become available which is important for establishing the utility of the SDQ for use in future epidemiological studies and clinical assessment. It is generally accepted that the evaluation of child behaviour could be culture-dependent (Heubeck, 2000) but since the political, social and economic structures of countries are quite similar, it can be expected that SDQ scores will be more similar across Scandinavian borders than in comparison with other countries. However, studies from both Norway and Finland report lower scale means than those found in the UK, and the crude descriptive comparisons presented in this paper suggest that differences across the Scandinavian countries may be rather small.

4. PUBLISHED ARTICLES

4.1 Validation of the parent and teacher SDQ in a clinical sample

A. Becker *, W. Woerner, M. Hasselhorn, T. Banaschewski, A. Rothenberger

Andreas Becker, Wolfgang Woerner, Tobias Banaschewski, Aribert Rothenberger

University of Göttingen, Child and Adolescent Psychiatry,

von-Siebold-Str. 5, D - 37075 Göttingen, Germany

Markus Hasselhorn

University of Göttingen, Georg-Elias-Müller Institute for Psychology,

Waldweg 26, D - **37073 Göttingen, Germany**

*Corresponding author: Andreas Becker

University of Göttingen

Child and Adolescent Psychiatry

von-Siebold-Str. 5

D - 37075 GöttingenGermany

Phone: ++49 - 551 - 39 2386

Fax: ++49 - 551 - 39 2696

Mail: abecker4@gwdg.de

Abstract

Objective This study examines whether the German translation of the Strengths and Difficulties Questionnaire (SDQ) is a reliable and valid screening instrument and whether it is as effective a tool for clinical diagnostics and scientific applications as the CBCL / TRF.

Methods We examined 543 children and adolescents (147 girls and 396 boys) with ages ranging from 5 to 17 years and correlated the results of the parent and teacher SDQ as well as the CBCL / TRF with clinical diagnoses. Furthermore, the adequacy of the scale structure of the SDQ was tested using confirmatory and exploratory factor analyses.

Results It was demonstrated that the scales of the parent and teacher versions were sufficiently homogeneous (.72 - .83). Correlations between SDQ scales and corresponding CBCL / TRF scales showed a high degree of congruence, while an exact replication of the original SDQ scale structure could also be achieved. Parent and teacher versions of both questionnaires presented with good validity, not only with regard to the discrimination between child psychiatric patients and a representative community sample, but also in the identification of different categories of disorders within the clinical sample.

Conclusion The parent and teacher SDQs proved to be valid and helpful questionnaires for use in the framework of a multi-dimensional behavioural assessment, and appear to be well-suited for screening purposes, longitudinal monitoring of therapeutic effects, and scientific research purposes.

Keywords: Strengths and Difficulties Questionnaire (SDQ), parent reports, teacher reports, screening instrument, validation

Introduction

The Strengths and Difficulties Questionnaire (SDQ,) is a questionnaire developed and thoroughly evaluated in Great Britain . It addresses negative and positive behavioural attributes of children and adolescents in the age range of 4 to 17 years and can be completed by parents or teachers and as a self-report by adolescents of age 11 or older . An extended version [7] includes a global rating of problems concerning mood, concentration, behaviour and interactions with others, and also addresses their impact and burden on others. Special follow-up versions are available for repeated administration (e.g., documentation of treatment effects). For non-commercial purposes, questionnaire forms can be downloaded from the internet (http://www.sdqinfo.com).

The items of the SDQ were originally selected following factor analyses of more extensive screening inventories and are based on nosological concepts. In accordance with the classification systems of DSM-IV [2] and ICD-10 [18], the items of the hyperactivity/inattention scale thus address overactivity (2 items), attention problems (2 items), and impulsivity (1 item). In contrast to similar questionnaires, special attention was given to a well-balanced inclusion of both positive and negative behavioural aspects, hence the raw scores of several items need to be inverted before they are summed up to scale values. Since the questionnaire can be completed in about 5 minutes, a positive effect on its acceptance by the responding informants can be expected, thus leading to low rates of refusal and missing answers.

Although the German version is frequently being used for clinical and scientific purposes, only few reports exist on its psychometric properties, such as normative [20] or validation studies [15].

After an initial multi-center validation study [15], the psychometric features and clinical results of the German parent version have been examined more comprehensively and systematically [3, 4]. The present report describes parent and teacher SDQ findings of a recent validation study in a large sample of child psychiatric patients.

First, several methods were used to evaluate the factor structure and construct validity of the German SDQ in this clinical setting. Next, we examined how well the SDQ subscales match with the corresponding scales of the German Child Behaviour Checklist (CBCL; [1, 5] or its teacher counterpart, the Teacher Report Form (TRF). Finally, it was investigated whether SDQ scores can predict the presence of a child psychiatric diagnosis (ICD-10) and detect specific diagnostical subgroups in a clinical sample to the same extent as the CBCL / TRF.

Methods

Sample and diagnostic subgroups

Between August 1998 and July 2000, parent, teacher, and self-reports of the German SDQ as well as the corresponding CBCL versions were collected from all outpatients and inpatients (age 5 to 17 years) of the University Clinic for Child and Adolescent Psychiatry in Göttingen.

To ensure that all analyses were based on an identical core sample with parent and teacher reports, 213 datasets were excluded as only the parent version of the questionnaires was available. A few additional records were discarded due to their large number of missing answers: 13 SDQs had more than two missing items on at least one of the five SDQ subscales and could not be used for the recommended

prorating of scale scores based on valid items, while 9 CBCLs or TRFs had to be discarded because over 20 item answers were missing.

Therefore, the results presented here are based on the remaining sets of data with completed parent and teacher ratings in the SDQ as well as the CBCL / TRF. This analysis sample consisted of 543 children and adolescents, including 147 girls with an average age of 10.8 ± 3.1 years, and 396 boys with an average age of 9.9 ± 2.8 years (Table 1).

--------------------------- please insert Table 1 here ---------------------------

After thorough clinical examination by child and adolescent psychiatrists, 380 of the 543 children and adolescents received a child psychiatric diagnosis on axis I (any diagnosis of psychiatric disorders, apart from categories F70 to F79, F80 to F83.99, F85 to F89, and F98). Most of the remaining 163 patients who were not considered to be psychopathologically disturbed presented with dyslexia or other specific learning disabilities.

To allow comparisons with other SDQ studies, patients' diagnoses were assigned to three diagnostical subcategories, following the same procedure as in previous reports (e.g., [15]) :

  1. Emotional disorders (F30 to F43.23, F43.25, F92.0 to F93.2, F93.8, F93.9)

  2. Oppositional / conduct disorders (F43.24, F43.25, F90.1, F91 to F92.99)

  3. Hyperactivity / attention-deficit disorders (F90 to F90.99)

Among the 380 patients with a relevant diagnosis on axis I, 28.6 % of the girls were diagnosed as having an emotional disorder, compared to 22.7 % of the boys. 27.3 %

of the boys had an oppositional / conduct disorder (girls: 14.3%), and 37.9% a disorder in the hyperactivity / attention-deficit category (girls: 15.6%). Table 1 gives a fuller description of the analysis sample.

Questionnaires

Behavioural problems of the children and adolescents were rated with the German versions of the SDQ as well as the German version of the CBCL [5]. All parents completed the parent SDQ and the CBCL, while all teachers completed the teacher SDQ and the TRF. The results of the four problem subscales of the SDQ (emotional symptoms, conduct problems, hyperactivity-inattention, and peer problems), the combined total difficulties score, and the positive scale assessing prosocial behaviour were computed, as were the corresponding CBCL and TRF scores for the attention problems, social problems, and anxious/depressed subscales, the composite scales for internalizing and externalizing problems, and the total problem score.

Statistical analysis

Statistical analyses were performed using SPSS (release 11.0). Due to the skewed distributions of the raw scale scores, most of the statistical calculations were carried out using non-parametric methods (Spearman rank correlations, Mann-Whitney U-tests). Goodness-of-fit estimates of the parent SDQ scale structure were obtained using Lisrel 8 software [14].

To evaluate and document the discriminative validity of the SDQ and CBCL scores, ROC analyses (Receiver Operating Characteristics; [12]) were calculated, which do not require a priori definition of a specific cut-off value for separating normal and abnormal scale scores. In ROC analyses, sensitivity (percentage of correctly

identified "cases") and specificity (percentage of correctly classified healthy "non-cases") are calculated for all possible cut-off points of a score, and then combined in a single value called "Area under the Curve" (AUC). This method allows a direct comparison of the predictive value of two different measures (e.g., SDQ and CBCL scores) with respect to a known outside criterion (e.g., clinically defined "cases" vs. "non-cases"). The obtained "Area under the Curve" reflects the discriminative validity: An AUC of .50 indicates chance discrimination, whereas an AUC value of 1.00 would reflect a perfect association between scale scores and clinical diagnosis.

Comparing the areas under two ROC curves computed for the same sample involves the calculation of a z ratio, derived from the two obtained AUC values, their standard errors, and the correlation between the two predictor scores [13].

Results

Scale means, scale homogeneity, and convergent validity

Table 2 shows the SDQ scale means found in the present clinical sample (parent and teacher scores) and in the German normative sample (parent version only; [19, 20]). In addition, the association between SDQ scores and the clinical status of the patients is also reported. As expected, presence of any clinical diagnosis of psychopathology is mirrored by significantly elevated scores in all SDQ scales, for both parent and teacher ratings. On the other hand, parent SDQ scores of patients without any child psychiatric diagnosis on axis I only show minor deviations from the respective scale means in the normative sample.

---------------------------- please insert table (2) here ----------------------------

Internal consistencies (Cronbach's α) obtained for adult informant-rated SDQ scales in this clinical sample are reported in Table 3. Although each subscale only consists of 5 items, none of the internal consistency coefficients was lower than .70 (.72 - .81 for parent subscales; .75 - .83 for teacher subscales). For the total difficulties score based on 20 items, parent- and teacher-rated instruments yielded identical coefficients (Cronbach's α = .83). Thus, both parent and teacher versions can be considered to be sufficiently reliable.

--------------------------- please insert table (3) here ---------------------------

To assess the convergent validity of the SDQ, correlations between SDQ and the corresponding CBCL / TRF scales were calculated (Table 3). A direct comparison of the results for the two adult informant versions is justified, since only those patients with completed parent and teacher ratings in both instruments were included in the analysis sample. As the CBCL does not contain a scale corresponding to the positive prosocial behaviour score of the SDQ, the association with the CBCL / TRF social problems scale was analysed instead.

All correlations between SDQ scales and their respective CBCL / TRF counterparts were highly significant, including both total problem scores as well as the specific subscales. In particular, the conduct problems subscale of the SDQ, consisting of 5 items, showed a very strong concurrence (.82 for parent ratings and .86 for teacher reports) with the CBCL composite score for externalizing symptoms (derived from 33 items in the parent version and 34 items in the TRF). Likewise, the 5-item SDQ subscale assessing emotional symptoms was also highly correlated (.77 and .80 for parent and teacher-rated versions) with the global CBCL and TRF scores for

internalizing symptoms (comprising 32 items in the CBCL and 36 items in the TRF). As expected, negative correlations (-.22 for parents and -.19 for the teacher version) were found between prosocial behaviour in the SDQ and the social problem scale of the CBCL / TRF. With respect to hyperactivity and attention problems, teacher judgements demonstrated higher agreement between the two compared instruments (.80) than parent ratings did (.64).

Evaluation of factor structure and factorial validity

Several analyses were carried out to verify the proposed 5-factor structure of the SDQ. The 25 parent-rated SDQ items were subjected to a confirmatory factor analysis (using Lisrel 8; [14]), which demonstrated a good model fit (adjusted goodness-of-fit index AGFI = .85) of the original 5-factor model and a sufficiently low error term (root-mean-square residual RMR = .07; for methodological details see [16]).

In addition, the factor structure of the German parent SDQ was also examined using an enlarged data set. This combined database was created by pooling the German standardization sample (N=930; [19, 20]) and the entire clinical validation sample with available parent SDQ data (N=756). Thus, a total of 1686 parent-rated SDQs from a mixed community and clinical sample were included in a principal component analysis with subsequent varimax rotation. The 5 extracted factors explained 53.9 % of the total variance. A high correspondence between the pattern of rotated loadings and the original SDQ scales was found. Exactly 5 of the 25 items had their highest loadings on each of the extracted factors, identically matching the intended scale structure.

A similar principal components analysis was performed with the teacher-rated SDQ, which had only been administered to the clinical sample. Here, the 5 extracted factors explained 57.9 % of the total variance, and a high degree of concordance of the pattern of rotated loadings with the original SDQ scales could again be shown.

---------------------------- please insert table (4) here ----------------------------

Further evidence of the factorial validity was sought by jointly entering 4 parent-rated SDQ subscales and the 4 corresponding CBCL scales in a combined principal components analysis with varimax rotation (Table 4). Together explaining 89.3 % of the total variance, each of the 4 extracted factors showed substantial loadings on the corresponding pair of scales from the two instruments.

Discriminative validity

In order to establish and compare the discriminative validity of the SDQ and the CBCL / TRF with respect to detecting "any psychiatric diagnosis" and different diagnostical subcategories, ROC analyses were carried out, with the respective scale scores serving as predictors of caseness (Table 5).

---------------------------- please insert table (5) here ----------------------------

Using the total difficulties score (SDQ) or total problems score (CBCL / TRF) to predict presence of any behavioural disorder, it could be shown that SDQ (parents: AUC=.771, teacher: AUC=.752) as well as CBCL (AUC=.775) and TRF (AUC=.760) were equally able to differentiate between patients with and without any clinical

diagnosis on axis I. Moreover, the examined subscales of the SDQ and CBCL / TRF could also effectively detect more specific clinical subgroups, again revealing only minor differences between the two inventories.

However, the hyperactivity-inattention scale of both the parent SDQ (AUC=.766) and the teacher SDQ (AUC=.792) allowed a more accurate prediction of the clinical subgroup with hyperactivity / attention-deficit disorders than the comparable attention problems scale of the CBCL (AUC=.703) and TRF (AUC=.722). The composite internalizing problems score of the parent CBCL was slightly superior (AUC=.733) to the corresponding scale of the SDQ (AUC=.691) in detecting patients with emotional disorders.

Comparisons of the two adult informants' judgements, also reported in Table 5, revealed approximately equal predictive values for the parent- and teacher-rated SDQ scales. None of the analysed SDQ scores showed a significantly superior prediction by either one of the two informants. This equivalence of predictive values of parent and teacher scales was similar when CBCL and TRF scores were regarded: Here, only the internalizing problems score in the parent-rated CBCL (AUC=.733) had a significantly higher predictive value with respect to emotional disorders than the teacher-rated TRF (AUC=.671).

Finally, the ability of the parent SDQ to distinguish between the present child psychiatric sample and the representative community sample used to standardize the German parent SDQ was also tested by ROC analysis of the combined parent-rated data available for a total of 1686 children and adolescents. The resulting AUC value of .829 for the total difficulties score showed a very high discriminative power of the SDQ to differentiate between subjects from the clinical and community samples.

Discussion

A comprehensive evaluation of the parent- and teacher-rated SDQ in a clinical child psychiatric sample showed that the German translation of this questionnaire corresponds well with the English original in regard to all investigated aspects. When examining the factor structure, the 25 items of the SDQ combined in exactly the proposed pattern. Thus, factor analysis of the German SDQ in a clinical sample further confirmed the original scale structure of the instrument. Additional analyses of the parent SDQ by means of confirmatory factor analysis showed a good model fit.

The demonstrated equivalence of the German SDQ with the CBCL / TRF is somewhat surprising, particularly since the SDQ only comprises about one fifth of the items included in the CBCL / TRF. In general, shorter scales are less reliable and valid than longer scales [17]. In this particular case, however, the brevity of the SDQ does not seem to negatively affect its validity.

Our findings show that there were only minor differences between SDQ and CBCL / TRF with respect to the discrimination between clinically defined subgroups with vs. without any behavioural disorders or representing specific diagnostical subcategories. However, the composite internalizing problems score of the CBCL was better able to detect emotional disorders than the 5-item emotional symptoms scale of the parent SDQ, while SDQ parent and teacher reports were superior in discriminating between patients with vs. without disorders in the hyperkinetic / attention-deficit subcategory. Very good discriminative power of the parent SDQ in distinguishing between community and clinical samples provided further evidence of its validity.

In summary, the results presented here confirm the usefulness of the SDQ for screening and epidemiological studies as well as for clinical research, and show that

the SDQ is not only a practical and economical, but also a valid tool for assessing different behavioural aspects of children and adolescents.

References

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

2. American Psychiatric Association (APA) (1994) Diagnostic and statistical manual of mental disorders. APA, Washington, DC

3. Banaschewski T, Woerner W, Becker A, Rothenberger A (2004) Diagnostik der Aufmerksamkeitsdefizitstörung: Unterstützung durch den Elternfragebogen zu Stärken und Schwächen des Kindes (SDQ). Monatszeitschrift für Kinderheilkunde, in press

4. Bettge S, Ravens-Sieberer U, Wietzker A, Holling H (2002) Ein Methodenvergleich der Child Behavior Checklist und des Strengths and Difficulties Questionnaire. Gesundheitswesen 64:119-124

5. Döpfner M, Melchers P, Fegert J, Lehmkuhl G, Lehmkuhl U, Schmeck K, Steinhausen H-C, Poustka F (1994) Deutschsprachige Konsensus-Versionen der Child Behavior Checklist (CBCL 4-18), der Teacher Report Form (TRF) und der Youth Self Report Form (YSR). Kindheit und Entwicklung 3:54-59

6. Goodman R (1997) The Strengths and Difficulties Questionnaire: a research note. J Child Psychol Psychiatry 38:581-586

7. Goodman R (1999) The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. J Child Psychol Psychiatry 40:791-801

8. Goodman R (2001) Psychometric properties of the Strengths and Difficulties Questionnaire. J Am Acad Child Adolesc Psychiatry 40:1337-1345

9. Goodman R, Meltzer H, Bailey V (1998) The Strengths and Difficulties Questionnaire: a pilot study on the validity of the self-report version. Eur Child Adolesc Psychiatry 7:125-130

10. Goodman R, Renfrew 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. Eur Child Adolesc Psychiatry 9:129-134

11. Goodman R, Scott S (1999) Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: is small beautiful? J Abnorm Child Psychol 27:17-24

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

13. Hanley JA, McNeil BJ (1983) A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 148:839-843

14. Jöreskog K, Sörbom D (1993) LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language. Lawrence Erlbaum Associates Publishers, Hillsdale, NJ

15. Klasen H, Woerner W, Wolke D, Meyer R, Overmeyer S, Kaschnitz W, Rothenberger A, Goodman R (2000) Comparing the German versions of the Strengths and Difficulties Questionnaire (SDQ-Deu) and the Child Behavior Checklist. Eur Child Adolesc Psychiatry 9:271-276

16. Stevens J (1996) Exploratory and confirmatory factor analysis. In: Stevens J (ed) Applied multivariate statistics for the social sciences. Lawrence Erlbaum Associates, Mawah, NJ, pp 362-427

17. Streiner DL, Norman GR (1989) PDQ statistics. C. V. Mosby, Scarborough

18. WHO - World Health Organization (1992) The ICD-10 classification of mental and behavioural disorders. Clinical descriptions and diagnostic guidelines. WHO, Geneva

19. Woerner W, Becker A, Rothenberger A (2004) Normative data and scale properties of the German parent SDQ. Eur Child Adolesc Psychiatry, supplement (this supplement)

20. Woerner W, Becker A, Friedrich C, Klasen H, Goodman R, Rothenberger A (2002) Normierung und Evaluation der deutschen Elternversion des Strengths and Difficulties Questionnaire (SDQ): Ergebnisse einer repräsentativen Felderhebung. Z Kinder Jugendpsychiatr Psychother 30:105-112

Table 1 : Göttingen clinical SDQ sample with completed parent and teacher ratings

Distribution of diagnostical categories and mean age (N=543)

Diagnostical categories	Boys (N=396)	Girls (N=147)	Total (N=543)
Any diagnosis on axis I	N=290 (73.2%)	N=90 (61.2%)	N=380 (70.0%)
Emotional disorders	N=90 (22.7%)	N=42 (28.6%)	N=132 (24.3%)
Conduct disorders	N=108 (27.3%)	N=21 (14.3%)	N=129 (23.8%)
Hyperactivity / Attention-deficit disorders	N=150 (37.9%)	N=23 (15.6%)	N=173 (31.9%)
mean age (years) (SD)	9.9 (2.8)	10.8 (3.1)	10.2 (2.9)

Table 2 :Scale means and correlations between parent- and teacher-reported SDQ
Normative data and clinical sample with (YES) or without (NO) any child psychiatric diagnosis (N=543)

SDQ scale	*parent SDQ*				*teacher SDQ*		
	norm	*clinic*			*clinic*		
	(N=930)	**(NO)** (N=163)	**(YES)** (N=380)	p	**(NO)** (N=163)	**(YES)** (N=380)	p
	Mean	**Mean**	**Mean**		**Mean**	**Mean**	
Total Difficulties	8.1	11.7	18.2	***	10.3	16.3	***
Emotional Symptoms	1.5	2.8	4.0	***	2.4	3.2	**
Conduct Problems	1.8	2.2	4.1	***	1.3	3.1	***
Hyperactivity-Inattention	3.2	4.6	6.4	***	4.1	6.0	***
Peer Problems	1.6	2.0	3.7	***	2.5	3.9	***
Prosocial Behaviour	7.6	7.7	6.6	***	6.6	4.9	***

*** $p \leq .001$; ** $p \leq .01$ (two-tailed Mann-Whitney U-Tests between patients with vs. without any child psychiatric diagnosis on axis I)

Table 3 : Reliability and validity coefficients of the German parent- and teacher-rated SDQ (N=543, Göttingen patients, 5 to 17 years)

SDQ and CBCL scales	Reliability		Validity	
	SDQ-P (Cronbach's α)	SDQ-T (Cronbach's α)	correlation SDQ-P/ CBCL	correlation SDQ-T/ TRF
SDQ - Total Difficulties **CBCL/TRF** - Total Problems	.83	.83	.83 ***	.87 ***
SDQ - Emotional Symptoms **CBCL/TRF** - Internalizing Problems	.72	.77	.77 ***	.80 ***
SDQ - Conduct Problems **CBCL/TRF** - Externalizing Problems	.73	.75	.82 ***	.86 ***
SDQ - Hyperactivity / Inattention **CBCL/TRF** - Attention Problems	.81	.83	.64 ***	.80 ***
SDQ - Peer Problems **CBCL/TRF** - Social Problems	.76	.77	.75 ***	.71 ***
SDQ - Prosocial Behaviour **CBCL/TRF** - Social Problems	.74	.81	-.22 ***	-.19 ***

*** $p \leq .001$ (Spearman rank correlations, two-tailed)

Table 4 : Factorial validity of the parent SDQ

(Göttingen patients, N=543; loadings ≥ .35 of the varimax-
rotated 4-factor solution, total explained variance = 89.3 %)

	Factor 1	Factor 2	Factor 3	Factor 4
Initial Eigenvalue :	4.14	1.48	0.91	0.61
Explained Variance (rotated) :	24.1%	23.1%	22.8%	19.3%
Peer Problems - **(SDQ)**	**.90**			
Social Problems - **(CBCL)**	**.87**			
Conduct Problems - **(SDQ)**		**.92**		
Externalizing Problems - **(CBCL)**		**.84**		
Emotional Symptoms - **(SDQ)**			**.94**	
Anxious/ Depressed - **(CBCL)**			**.85**	
Hyperactivity-Inattention - **(SDQ)**				**.89**
Attention Problems - **(CBCL)**	.42			**.71**

Table 5 : Prediction of clinical diagnosis by parent- and teacher-rated SDQ or CBCL/TRF scales

(AUC = cut-off-independent area under the ROC curve; Göttingen patients; N=543 with parent <u>and</u> teacher ratings,
including N=163 without any diagnosis on axis I)

Predicted clinical target group	Considered questionnaire scale	Parent		Teacher		P vs. T
		AUC	p[1]	AUC	p[1]	p[2]
Any diagnosis on axis I yes: N=380 vs. no: N=163	**SDQ** - Total Difficulties	.771	ns	.752	ns	ns
	CBCL/TRF - Total Problems	.775		.760		ns
Emotional disorders yes: N=132 vs. no: N=411	**SDQ** - Emotional Symptoms	.691	*	.648	ns	ns
	CBCL/TRF - Internalizing Problems	.733		.671		*
Oppositional / conduct disorders yes: N=129 vs. no: N=414	**SDQ** - Conduct Problems	.813	ns	.821	ns	ns
	CBCL/TRF - Externalizing Problems	.838		.828		ns
Hyperactivity / Attention-deficit disorders yes: N=173 vs. no: N=370	**SDQ** - Hyperactivity / Inattention	.766	***	.792	***	ns
	CBCL/TRF - Attention Problems	.703		.722		ns

*** p ≤ .001; * p ≤ .05; ns: not significant

[1] two-tailed z-test comparing AUC differences between <u>instruments</u> (parent SDQ vs. CBCL and teacher-reported SDQ vs. TRF)

[2] two-tailed z-test comparing AUC differences between <u>informants</u> (parent vs. teacher SDQ and CBCL vs. TRF)

4.2 Evaluation of the self-reported SDQ in a clinical setting: Do self-reports tell us more than ratings by adult informants?

A. Becker*, N. Hagenberg, V. Roessner, W. Woerner, A. Rothenberger

Andreas Becker, Nicola Hagenberg, Veit Roessner, Wolfgang Woerner, Aribert Rothenberger
University of Göttingen, Child and Adolescent Psychiatry,
von-Siebold-Str. 5, D - 37075 Göttingen, Germany

*Corresponding author: Andreas Becker
                          University of Göttingen
                          Child and Adolescent Psychiatry
                          von-Siebold-Str. 5
                          D - 37075 Göttingen Germany
                          Phone: ++49 - 551 - 39 2386
                          Fax: ++49 - 551 - 39 2696
                          Mail: abecker4@gwdg.de

Abstract

Objective The aim of this study was to evaluate the German self-reported Strengths and Difficulties Questionnaire (SDQ) in a clinical setting. We also investigated whether this additional information gathered directly from older children and adolescents improves the prediction of clinical status when external ratings from their parents and/or teachers are already available.

Methods SDQ self-reports were collected from 214 in- and outpatients (81 girls and 133 boys) aged 11 to 17 years who were seen at the department of child and adolescent psychiatry of the University of Göttingen. Results obtained with the self-rated questionnaire were compared with the parent and teacher SDQs, corresponding CBCL / YSR scores, and the clinical diagnostic classification. Finally, the additional diagnostical benefits of the self-reports were examined.

Results The scales of the SDQ self-report proved to be sufficiently homogeneous, and acceptable correlations were found with the equivalent parent and teacher ratings. The self-rated version of the SDQ demonstrated good validity with respect to the differentiation between clinically defined cases and non-cases and in detecting various subcategories of psychiatric disorders within the clinic sample. SDQ self-reports significantly contributed to the prediction of diagnostic status, specifically if only parent <u>or</u> teacher ratings were available.

Conclusions The self-rated version of the SDQ was shown to be a reliable and valid method for the assessment of behavioural problems in children and adolescents. In the absence of adult informant reports from parents and teachers, the diagnostical value of self-ratings was also demonstrated.

Keywords: Strengths and Difficulties Questionnaire (SDQ), self-report, YSR, ROC analysis

Introduction

Various studies have shown that a considerable percentage of children and adolescents meet the criteria for emotional or behavioural disorders [9, 26]. Although their problems can create much distress for daily life and well-being, many adolescents with psychiatric disorders receive no professional help, even though early intervention could prevent persistence of the disorder into adult life [16].

Evaluation of child psychiatric disturbances is mainly based on clinical interviews with parents and teachers, assessment of problem behaviour with various questionnaires, and on observations of behaviour in a diagnostic setting. For a comprehensive evaluation of such disturbances it is necessary to draw on information from the child or adolescent him/herself as a valuable source for describing the patient's feelings, moods, and subjective experiences [32].

For this purpose, structured interviews such as the Diagnostic Interview of Psychological Disorders in children and adolescents (German child DIPS [29]), the Kiddie-SADS interview [18] or the NIMH Diagnostic Interview Schedule for Children (DISC [7]) can be employed. In these assessment instruments, the interview questions are standardized to reduce variance of both applied criteria and collected information, thus increasing reliability and allowing a better comparison between patients or studies. However, these methods are mainly applied in the framework of scientific investigations. Due to the considerable amount of time required as well as the necessary degree of clinical expertise of the interviewer, structured interviews are of limited importance for primary health care in a clinical setting [17]. Given these problems and limitations, child and adolescent psychiatry has benefitted from the development and application of screening questionnaires which are shorter, less costly, and easier to apply [10]. Using such tools for gathering information, the

clinician can quickly get an impression of the most relevant psychopathological symptoms in children or adolescents.

Since younger children only have a limited ability for introspection and are often unable to judge and report on their emotions or behaviour, their direct assessment is considered to be only of limited diagnostical value [3]. Children under the age of 11 years are also often unable to transfer their daily experiences onto a questionnaire made up of only few items. Thus, it seems reasonable to only utilize the information from self-reports of children aged about 11 years or older.

Several studies have shown that the act of seeking clinical assistance has considerable influence on the validity of the characteristics being assessed. For example, a distinct answer tendency in a non-clinical random sample showed that children and adolescents described themselves as having more behavioural problems than were reported by their parents and teachers [27]. The opposite effect was demonstrated in a clinical sample [14], where dissimulation tendencies made children and adolescents under-report their behavioural problems.

Methods and results when examining the utility of self-reports and their contribution to clinical diagnostics were found to be very heterogeneous. One approach to assess the quality of the information obtained with the self-rated version of the SDQ involves comparisons with the well-evaluated Youth Self-Report (YSR; [2]). The YSR is the self-report derivative of the Child Behavior Checklist (CBCL; [1]), and has been widely implemented to describe relevant problems in children and adolescents [6]. Initial reports on the internal consistency of composite scales of the YSR [2] were later replicated in Switzerland [28], Norway [22], and Germany [25]. Cross-informant

agreement between CBCL parent reports and YSR self-ratings usually yielded only moderate correlations (e.g., [19, 30]). Nevertheless, Morgan and coworkers [23] successfully predicted clinical (DSM-III R) diagnosis on the basis of YSR subscales.

However, it must be noted that the scale structure of the YSR was empirically determined, and only some of its problem scales correspond to specific diagnostical categories. For example, although several symptoms of attention-deficit and hyperactivity disorders are included in the attention problems subscale of the YSR, the items on this scale do not cover the entire range of criteria specified for the corresponding clinical definitions according to DSM-IV or ICD-10. In consequence, validity of this YSR subscale assessing attention disorders was only partly confirmed in a clinical sample [8].

On the other hand, the Strengths and Difficulties Questionnaire (SDQ) was designed by Goodman [11] on the basis of nosological criteria (ICD-10). Like its parent- and teacher-rated counterparts, the self-report version of the SDQ only comprises 25 items and addresses a well-balanced number of positive and negative behavioural aspects, in contrast to the considerably longer YSR. In Germany, normative data have only been established for the parent-rated version of the SDQ ([31]; see also Woerner et al., in this supplement volume).

Although the SDQ has been used as a diagnostical instrument only since 1997, studies from several countries have already reported psychometric properties of its self-rated version. Results on the internal consistency (Cronbach's α) demonstrate the homogeneous scale structure, with reliabilites for the total difficulties score ranging between 0.82 (England: [13]), 0.78 (Netherlands [24]), and 0.71 (Finland:

[20]). Inspite of their shortness, good respective results were also obtained for the five self-rated subscales.

Analyses of cross-informant agreement yielded a similar correspondence between the self-rated SDQ and the parent version as had been found for the YSR: Correlations for the total difficulties score of the SDQ obtained in clinical or community-based sample in the Netherlands, in Finland, and in the UK ranged from 0.40 to 0.46.

A review of studies with structured diagnostic interviews [17] demonstrated that the level of agreement between self-reports and parent ratings depends on the type of disorder. Similar findings were noted with the SDQ in a clinical random sample [13], with a higher degree of correspondence between parent and self-rating for externalizing disorders and lower cross-informant agreement for internalizing disorders. However, the opposite tendency was observed in community samples [12, 24].

Heterogeneous results were also reported with respect to mean scale scores of adult informants and self-reports. In an English clinical sample, Goodman and colleagues [13] found that older children and adolescents often rated themselves in the normal or borderline range of the total difficulties score, while their parents or teachers usually reported a higher degree of impairment. In contrast, Finnish children and adolescents described themselves as having considerably more behavioural problems than their parents did [20].

Consistent gender differences in self-rated SDQ scores were documented in a large number of reports. For example, Dutch girls had significantly higher values on the emotional symptoms and prosocial behaviour subscale, while boys scored higher on

the conduct problems scale [24]. Similar findings were reported from Finland [21], where girls in their 7th or 9th school year scored much higher than boys on emotional symptoms and prosocial behaviour, but also had a higher total difficulties score. In this Finnish sample, boys again reported more conduct problems than girls.

To evaluate the validity of the self-rated SDQ, it was compared with the corresponding YSR scores. In Finland, this yielded correlations of 0.71 for the total difficulties score and between 0.43 and 0.68 for the subscales. In a German non-clinical random sample, Bettge and coworkers [5] obtained a somewhat lower correlation of 0.58 for the total difficulties score.

Based on these previous results gathered for the self-rated SDQ, the present study comprehensively evaluates the German version in a clinical setting. We examined how well the information from children and adolescents matches with the judgements by their parents and teachers. To further analyze the psychometric properties of the SDQ self-report scales, these were compared with the corresponding YSR scores and with the clinical diagnosis as determined according to ICD-10 criteria. Here, it was examined how accurately the self-report version of the SDQ is able to discriminate between older children with or without any diagnosis of a child psychiatric disorder. At a more detailed level, SDQ scores were used to detect more specific diagnostical subgroups. Finally, another question of particular interest was whether the self-ratings gathered from children and adolescents additionally improve the prediction of their clinical diagnosis when external ratings from parents and/or teachers are also available.

Methods

Clinical sample and diagnostical subgroups

In the present study, data were collected from in- and outpatients of the Child and Adolescent Psychiatry department of the University of Göttingen and from their parents and teachers. At the time of their assessment (August 1998 until July 2000) all patients were between 11 and 17 years old. SDQ self-reports and parent-rated SDQs were available for all participants, as well as self- and parent-rated YSR / CBCL scores. Independently determined clinical diagnoses by board-certified child and adolescent psychiatrists were based on ICD-10 criteria. The present analysis sample with complete self-report and parent data numbers 214 older children and adolescents, including 133 boys (62%) with a mean age of 13.8 years (SD: 1.9) and 81 girls (38%) with a mean age of 14.8 years (SD: 1.6).

------------------------------ please insert Table 1 here --------------------------

As shown in Table 1, 194 patients (90.7% of the entire analysis sample) received at least one clinical diagnosis of a child psychiatric disorder (any diagnosis on axis I except F70 to F79, F80 to F83.99, F85 to F89, and F98). Subdividing into broad-band diagnostical subgroups, 100 children and adolescents (46.7%) showed an emotional disorder; 69 (32.2%) had an oppositional/conduct disorder, and 43 patients (20.1%) were found in the subcategory of hyperactivity/attention-deficit disorders (see for a more detailed definition of diagnostical subgroups).

In order to compare the self-report results with those obtained with parent and teacher ratings, some of the analyses were carried out for a subsample of 124 patients who had complete sets of SDQ data from all three informants (i.e., parent, teacher, and self-reports).

Statistical analysis

For all three versions of the SDQ, means and reliabilities (Cronbach's α) were calculated for the total difficulties score and each of the five subscales. Spearman rank correlations between the patients' age and SDQ scores were also determined. To examine the factor structure of the SDQ self-report, a principal component analysis with varimax rotation was carried out. Rank correlations between the self-reports and adult-rated scores were calculated for the entire analysis sample and separately for each gender subgroup.

Since normative data for the German SDQ self-report were not available at the time of writing, evaluation in this study was performed without applying predefined cut-off values. Hence, cut-off-independent Receiver Operating Characteristics (ROC) analyses were chosen to quantify and compare the discriminative validity of the SDQ. ROC analyses yield a cut-off-independent measure called AUC (area under the curve), which summarizes the sensitivity and specificity of all possible cut-off positions with respect to a known external criterion (e.g., clinical diagnosis of hyperactivity/attention-deficit disorder: yes vs. no). The obtained AUC values reflect the discriminative validity, and allow direct comparisons between the predictive value of two different instruments. An AUC of 1.0 would mean that the instrument discriminates perfectly between children with and without a certain diagnosis, an AUC of 0.5 is obtained when a measure merely discriminates at chance level. For the present study, ROC analyses were used to compare differences between both self-reports (self-rated SDQ vs. YSR), between the two parent questionnaires (parent SDQ vs. CBCL), and between self-reports and parent reports using the same instrument (self-rated vs. parent SDQ; YSR vs. CBCL). The algorithm used for

statistical comparisons between two obtained AUC values has been described in detail by Hanley & McNeil [15].

Furthermore, adjusted R² values obtained in blockwise multiple regression analyses were examined, to show the effects of different combinations of SDQ informants on the explained variance of the dependent target variable (i.e., presence of any psychiatric disorder or a specific clinical diagnosis).

Results

Scale means, scale homogeneity and correlations with age

For each of the SDQ versions, Table 2 reports mean scale scores, the homogeneity coefficient of each scale, and the correlations between the patients' age and their SDQ scores.

Examination of the pattern of scale means shows that the children and adolescents in this clinical sample described themselves as having fewer problems and showing more prosocial behaviour than according to their parents' reports. In contrast, teacher ratings of their pupils mentioned fewer emotional symptoms and less prosocial behaviour.

Analysis of the reliabilities (Cronbach's α) of the SDQ scales yielded satisfactory results (total difficulties score in self-reports: $\alpha= 0.78$; compared to $\alpha= 0.82$ for parent and teacher ratings, respectively). Although only containing five items each, subscales of the self-reported SDQ were also shown to be acceptably homogeneous (e.g., emotional symptoms: $\alpha= 0.77$). Only the internal consistency of the conduct problems subscale fell below 0.60. Subscales measuring conduct problems,

hyperactivity-inattention, and peer problems were clearly more homogeneous in parent and teacher ratings than in the self-reports (Table 2).

------------------------------ please insert Table 2 here --------------------------

For all three informants, correlations between the patients' age and SDQ scores showed a significant negative age effect on the conduct problems subscale (self-report: -.19). While parents as well as teachers reported significantly fewer total difficulties and less hyperactive/inattentive behaviour with increasing age of the children, neither of these two scores was strongly related to age when obtained from the patients themselves. Conversely, older patients scored higher on self-reported emotional symptoms and prosocial behaviour than younger ones, whereas the same scores taken from parent and teacher ratings reflected either much weaker age effects or none at all.

Evaluation of the factorial structure

The factor structure of the self-rated SDQ within this clinic sample was examined by carrying out a principal component analysis with varimax rotation. The five extracted factors explained 51.4% of the total variance, and the resulting pattern of loadings showed a high conformance with the original SDQ scales.

Evaluation of cross-informant agreement

Agreement between self-reports and adult informant-rated scores was determined by calculating rank correlations between corresponding scales (Table 3). Although all coefficients attained statistical significance, it was noted that the total difficulties score and the emotional symptoms subscale showed a much weaker cross-informant

agreement than the other SDQ scores, particularly when compared to the conduct problems and peer problems subscales. Overall, self-ratings were more closely associated with their parents' reports than with those by their teachers.

--------------------------- please insert Table 3 here ---------------------------

When cross-informant agreement was regarded separately in male and female subsamples, correlations between self-reports and parent ratings were found to be much higher for girls than for boys. Girls' self-ratings showed significantly higher agreement with their parents' judgements in the total difficulties score (boys: 0.29; girls: 0.61) and with respect to hyperactive-inattention, conduct problems, and peer problems. The subscales assessing emotional symptoms and prosocial behaviour did not show such gender-dependent effects, nor did any of the correlations between self-reports and teacher ratings (Table 3).

Evaluation of discriminative validity

ROC analyses were carried out to evaluate the discriminative properties of the self-rated SDQ in comparison to the YSR. The predictive power of the self-report measures was also compared to parent reports (parent SDQ and CBCL). Both the ability to distinguish between patients with and without any psychiatric diagnosis and detection of diagnostical subgroups were examined in these analyses (Table 4).

Using either their total problem scores or their specific subscales, both SDQ versions (self-report or parent-rated) were similarly effective in predicting the respective clinical categories as the YSR or CBCL. Thus, the total difficulties score of the SDQ self-report was just as able to distinguish between patients with vs. without any psychiatric diagnosis (AUC = .835) as was the YSR total problems score (AUC = .810), as shown in the left column of Table 4. Moreover, it was seen that the

subscales of the self-rated SDQ could effectively detect clinical subgroups of patients, particularly those with oppositional/conduct disorders. Again, there were only very small differences between the AUC measures obtained with the two self-report instruments, demonstrating a similar diagnostical efficiency.

---------------------------- please insert Table 4 here ----------------------------

Comparison between the predictive values of self-reports and parent ratings revealed a slight superiority of the self-rated SDQ, since its total difficulties score was marginally better at predicting the presence of any psychiatric disorders than the corresponding overall total score obtained from parent reports. Parent ratings, on the other hand, seemed to be slightly better predictors of more specifically defined diagnostical subgroups than self-reports, particularly concerning patients with clinically diagnosed hyperactivity and inattention disorders. This tendency was observed for both SDQ and YSR / CBCL instruments (right column of Table 4).

Predictive value of different combinations of SDQ informants

A different methodological approach, used to evaluate the relative contribution of individual SDQ informants to the overall prediction of clinical diagnoses, involved comparisons of adjusted R^2 values derived from regression analyses. Each adjusted R^2 shown in Table 5 reflects the proportion of variance of the dependent variable (i.e., subsamples with or without the indicated clinical diagnosis) that can be predicted using SDQ information from one or more sources.

---------------------------- please insert Table 5 here ----------------------------

As shown in Table 5, information from parents and teachers as well as from the children and adolescents themselves were all helpful in predicting a patient's clinical status, although the overall accuracy of predicting presence vs. absence of any or more specific clinical diagnosis was rather low within this clinical sample of older children and adolescents. When considering only ratings from one single source, SDQ self-reports proved to be as good as the parent-rated version at predicting the overall presence of a psychiatric diagnosis. However, using only parent or teacher information to detect specific diagnostical subgroups, each of these sources yielded better results than self-ratings alone - the exception being the emotional disorder subcategory, which was best predicted by the parent-rated emotional symptoms subscale and less well detected by teachers.

When, in addition to parent reports, SDQ self-ratings were added as predictor, the proportion of explained variance increased significantly (e.g., for conduct/oppositional disorders: parent-only = 24 %, parent + self-report = 28 %). Similar results were found when adding the self-report to teacher ratings (emotional symptoms: teacher-only = 7 %, teacher + self-report = 12 %). However, using information from parent and teacher ratings as predictors (i.e., combining data from both adult informants), the percentage of explained variance was generally higher than combining self-reported information with either parent or teacher ratings (e.g., for conduct/oppositional disorders: parent + self-report = 28 %, teacher + self-report = 32 %, and parent + teacher = 35 %). While the combination of all three sources (i.e., using self-ratings in addition to information from both teacher- and parent-rated SDQs) yielded an additional slight increase in adjusted R², none of these observed increments reached statistical significance.

Thus, it was demonstrated that the ability to predict the indicated diagnostical subgroups was significantly improved when the self-report version was added to either parent <u>or</u> teacher ratings. If information from both adult sources (i.e., parents <u>and</u> teachers) was already available, the additional inclusion of self-ratings did not provide any improvement in predictive power.

Discussion

The self-report version of the German SDQ was evaluated in a clinical sample of 214 in- and outpatients, aiming to determine the psychometric properties of this self-rated questionnaire and to investigate the diagnostical utility of adding self-reported information to external ratings by parents and teachers. Evaluation of the homogeneity of the total score and individual subscales demonstrated satisfactory reliability. In comparison with the corresponding scales of the parent version, SDQ self-ratings showed higher reliability for the emotional symptoms subscale, while, in contrast, other scales (conduct problems, hyperactivity-inattention, and peer problems) were more homogeneous in parent ratings.

Among the investigated patients, age effects were observed for some subscales of the SDQ self-report, older adolescents reporting more emotional symptoms and more prosocial behaviour than younger children. In line with the results of a previous clinical study [13], patients in the present clinical sample rated themselves as having fewer behavioural problems than according to their parents' reports.

Cross-informant correlations of self-ratings were found to be higher with parent than with teacher reports, replicating findings from other studies (e.g., [3]). Compared to boys, self-reports of girls showed considerably stronger agreement with their parents concerning peer relations, hyperactivity, inattention, and conduct problems.

The self-rated SDQ also turned out to possess at least comparable discriminative validity as the well-established but much longer YSR, both questionnaires being equally able to differentiate between patients with and without child psychiatric diagnoses. Further comparisons between the SDQ self-report and the corresponding YSR scales failed to find substantial differences in effectively detecting specific diagnostical subgroups.

SDQ self-reports were slightly better than the parent SDQ at identifying subgroups with and without any clinical diagnosis, while the parent version allowed a more accurate prediction of hyperactivity and attention-deficit disorders. Similar results have previously been reported by Klasen and coworkers [19].

The potential benefits of additionally considering self-report questionnaires to improve prediction of clinical status were investigated using a multiple regression approach. Here, SDQ scores as reported by either one, two, or all three informants served to predict presence of any child psychiatric diagnosis at all and, at a more specific level, broader-band diagnostical subcategories. When self-report information was added to either parent or teacher ratings, prediction of the clinical status improved. If, however, information from both other sources (i.e., parents and teachers) was already available, the additional inclusion of self-report data was unable to provide any further predictive power. It should be noted that, within this clinical sample, overall prediction of presence vs. absence of psychiatric diagnoses was rather poor, presumably reflecting an abundance of behavioural problems even in the fewer cases who did not fully meet the criteria for any or a specific clinical diagnosis.

In conclusion, the present results demonstrate that SDQ self-ratings from children and adolescents represent a useful contribution to the diagnostical process.

Comparing the described findings obtained for the German SDQ self-report with those from other countries, our results on scale means, internal consistency, cross-informant correlations, and discriminative validity agree well with reports from England [13], the Netherlands [24], and Finland [20], thus further establishing the self-rated version of this easily handled and economical instrument as a powerful and reliable tool for a multitude of clinical and research applications.

References

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

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

3.   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

4.   Becker A, Woerner W, Hasselhorn M, Banaschewski T, Rothenberger A (2004) Validation of the parent and teacher SDQ in a clinical sample. European Child & Adolescent Psychiatry, submitted (this supplement)

5.   Bettge S, Ravens-Sieberer U, Wietzker A, Holling H (2002) Ein Methodenvergleich der Child Behavior Checklist und des Strengths and Difficulties Questionnaire. Gesundheitswesen 64:119-124

6.   Broberg AG, Ekeroth K, Gustafsson PA, Hansson K, Hägglöf B, Ivarsson T, Larsson B (2001) Self-reported competencies and problems among Swedish adolescents: a normative study of the YSR. European Child & Adolescent Psychiatry 10:186-193

7.   Costello AJ, Edelbrook LS, Dulcan MK, Kalas R, Klaric SH (1984) Report on the NIMH Diagnostic Interview Schedule for Children (DISC). National Institute of Mental Health, Washington, DC

8.   Döpfner M, Berner W, Lehmkuhl G (1995) Reliabilität und faktorielle Validität des Youth Self-Reports der Child-Behavior Checklist bei einer klinischen Stichprobe. Diagnostica 41:221-244

9.   Fergusson DM, Horwood LJ, Lynskey MT (1993) Prevalence and comorbidity of DSM-III-R diagnoses in a birth cohort of 15 year olds. Journal of the American Academy of Child and Adolescent Psychiatry 32:1127-1134

10.   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 32:677-693

11.   Goodman R (1997) The Strengths and Difficulties Questionnaire: a research note. Journal of Child Psychology and Psychiatry 38:581-586

12.   Goodman R, Ford T, Simmons H, Gatward R, Meltzer H (2000) Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. British Journal of Psychiatry 177:534-539

13.   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 & Adolescent Psychiatry 7:125-130

14.   Görtz A, Döpfner M, Nowak A, Bonus B, Lehmkuhl G (2002) Are self-reports of adolescents helpful in the assessment of Attention Deficit/Hyperactivity Disorders? An analysis with the assessment system DISYPS. Kindheit und Entwicklung 11:82-89

15.   Hanley JA, McNeil BJ (1983) A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 148:839-843

16.   Harrington R, Rutter M, Fombonne E (1996) Developmental pathways in depression: multiple meanings, antecedents, and endpoints. Development and Psychopathology 8: 601-616

17.   Hodges K (1993) Structured interviews for assessing children. Journal of Child Psychology and Psychiatry 34:49-68

18.   Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, Williamson D, Ryan N (1997) Schedule for Affective Disorders and Schizophrenia for School-Age Children - Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. Journal of the American Academy of Child and Adolescent Psychiatry 36:980-988

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

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

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

22.   Kvernmo S, Heyerdahl S (1998) Influence of ethnic factors on behavior problems in indigenous Sami and majority Norwegian adolescents. Journal of the American Academy of Child and Adolescent Psychiatry 37:743-751

23.   Morgan CJ, Cauce AM (1999) Predicting DSM-III-R disorders from the Youth Self-Report: analysis of data from a field study. Journal of the American Academy of Child and Adolescent Psychiatry 38:1237-1245

24.   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 & Adolescent Psychiatry 12:1-8

25. Roth M (2000) Evaluation of the Youth Self-Report in a nonclinical sample. Zeitschrift für Differentielle und Diagnostische Psychologie 21:105-110

26. Saunders SM, Resnick MD, Hoberman HM, Blum RW (1994) Formal help-seeking behavior of adolescents identifying themselves as having mental health problems. Journal of the American Academy of Child and Adolescent Psychiatry 33:718-728

27. Sawyer MG, Baghurst P, Mathias J (1992) Differences between informants' reports describing emotional and behavioural problems in community and clinic-referred children: a research note. Journal of Child Psychology and Psychiatry 33:441-449

28. Steinhausen H-C, Winkler-Metzke C (1998) Youth Self-Report of behavioral and emotional problems in a Swiss epidemiological study. Journal of Youth and Adolescence 27:429-441

29. Unnewehr S, Schneider S, Margraf J (1995) Diagnostisches Interview psychischer Störungen im Kindes- und Jugendalter. Springer, Berlin

30. Verhulst FC, van der Ende J (1992) Agreement between parents' reports and adolescents' self-reports of problem behavior. Journal of Child Psychology and Psychiatry 33:1011-1023

31. Woerner W, Becker A, Friedrich C, Klasen H, Goodman R, Rothenberger A (2002) Normierung und Evaluation der deutschen Elternversion des Strengths and Difficulties Questionnaire (SDQ): Ergebnisse einer repräsentativen Felderhebung. Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie 30:105-112

32. Zwaanswijk M, van der Ende J, Verhaak PF, Bensing JM, Verhulst FC (2003) Factors associated with adolescent mental health service need and utilization. Journal of the American Academy of Child and Adolescent Psychiatry 42:692-700

Table 1: Clinical sample with completed self-report and parent SDQ

Distribution of diagnostical categories and mean age (N=214)

Diagnostical categories	Boys N=133	Girls N=81	Total N=214
Any diagnosis on axis I	N=121 (91.0%)	N=73 (90.1%)	N=194 (90.7%)
Emotional disorders	N=55 (41.4%)	N=45 (55.6%)	N=100 (46.7%)
Conduct / oppositional disorders	N=49 (36.8%)	N=20 (24.7%)	N=69 (32.2%)
Hyperactivity / attention-deficit disorders	N=40 (30.1%)	N=3 (3.7%)	N=43 (20.1%)
Mean age (years) (SD)	13.8 (1.9)	14.8 (1.6)	14.2 (1.9)

Table 2: Scale evaluation of the German parent, teacher, and self-report SDQ in a clinical sample

(11- to 17-year-old Göttingen in- and outpatients, N=124 with complete data from all three informants)

SDQ scale	Scale properties								
	self-report			*parent rating*			*teacher rating*		
	scale mean X	*reliability* [1] α	*r with age*	*scale mean* X	*reliability* [1] α	*r with age*	*scale mean* X	*reliability* [1] α	*r with age*
Total Difficulties Score	15.0	.78	.01	17.0	.82	-.16*	14.7	.82	-.20*
Emotional Symptoms	3.8	.77	.31***	4.4	.73	.12	3.0	.77	.15*
Conduct Problems	3.0	.58	-.19*	4.0	.79	-.21*	2.8	.76	-.28***
Hyperactivity / Inattention	4.9	.65	-.14	5.8	.82	-.28***	5.3	.87	-.25**
Peer Problems	3.2	.65	-.01	3.4	.76	-.09	3.6	.75	-.08
Prosocial Behaviour	6.9	.78	.24**	6.0	.76	.07	4.5	.82	.18*

* $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$ (Spearman rank correlations, one-tailed)

[1] homogeneity coefficient (Cronbach's α)

Table 3: **Correlations between SDQ scales from different informants**

(11- to 17-year-old Göttingen in- and outpatients with available parent, teacher, and self-report SDQs, N=124)

SDQ scale	self x parent report [1]	male [1] (N=83)	female [1] (N=41)	p [2]	self x teacher [1] report	male [1] (N=83)	female [1] (N=41)	p [2]
Total Difficulties	.39	.29	*.61*	*	.27	.21	*.42*	*ns*
Emotional Symptoms	.30	.31	.27	ns	.29	.20	.30	ns
Conduct Problems	.51	.42	.66	*	.50	.48	.59	ns
Hyperactivity / Inattention	.44	.33	.62	*	.38	.31	.47	ns
Peer Problems	.57	.49	.70	*	.46	.48	.39	ns
Prosocial Behaviour	.46	.37	.60	ns	.32	.21	.41	ns

[1] all correlation coefficients significant at $p \leq .05$ (one-tailed Spearman rank correlations)

[2] * $p \leq .05$; ns: not significant (one-tailed z-test comparing correlation coefficients obtained for male vs. female subgroups)

Table 4: Prediction of clinical diagnosis by parent- and self-rated SDQ or CBCL / YSR scales

(AUC : Area under the curve = cut-off-independent area under the ROC curve; N=214 Göttingen in- and outpatients with parent reports and self-ratings, including N=20 without any diagnosis on axis I)

Predicted clinical target group	Considered questionnaire scale	Self-report		Parent		
		AUC	p^1	AUC	p^1	p^2
Any diagnosis on axis I	**SDQ** - Total Difficulties Score	.835		.738		т
yes: N=194 vs. no: N=20	**YSR / CBCL** - Total Problems Score	.810	ns	.746	ns	ns
Emotional disorders	**SDQ** - Emotional Symptoms	.690		.712		ns
yes: N=100 vs. no: N=114	**YSR / CBCL** - Internalizing Problems	.698	ns	.728	ns	ns
Oppositional / conduct disorders	**SDQ** - Conduct Problems	.773		.824		ns
yes: N=69 vs. no: N=145	**YSR / CBCL** - Externalizing Problems	.742	ns	.832	ns	**
Hyperactivity / attention-deficit disorders	**SDQ** - Hyperactivity / Inattention	.684		.759	т	т
yes: N=43 vs. no: N=171	**YSR / CBCL** - Attention Problems	.623	ns	.707		т

** p ≤ .01; т p ≤ .10; ns: not significant

[1] two-tailed z-test comparing AUC differences between <u>instruments</u> (parent SDQ vs. CBCL ; self-rated SDQ vs. YSR)

[2] two-tailed z-test comparing AUC differences between <u>informants</u> (parent- vs. self-rated SDQ ; CBCL vs. YSR)

Table 5 :   Prediction of diagnostical categories using different combinations of SDQ informants

(adjusted R² values representing % of predicted variance of the respective target category; core sample of 11- to 17-year-old Göttingen in- and outpatients with complete SDQ data from all three informants; N = 124)

Diagnostical categories / scales	self only	parent only	parent + self	p	teacher only	teacher + self	p	parent + teacher	parent + teacher + self	p
Any diagnosis on axis I (N= 112 vs. 12) predicted by: Total difficulties score (yes vs. no)	.10	.09	.12	*	.13	.17	*	.15	.17	ns
Emotional disorders (N= 55 vs. 69) predicted by: Emotional symptoms scale (yes vs. no)	.09	.17	.20	*	.07	.12	*	.20	.21	ns
Conduct / oppositional disorders (N= 46 vs. 78) predicted by: Conduct problems scale (yes vs. no)	.18	.24	.28	*	.30	.32	*	.35	.36	ns
Hyperactivity / attention-deficit (N= 28 vs. 96) predicted by: Hyperactivity / Inattention scale (yes vs. no)	.09	.12	.14	*	.12	.15	*	.16	.17	ns

* p ≤ .05; ns: not significant   (two-tailed z-test of the effect of adding self-report data to parent and/or teacher ratings)

4.3 Anwendung des SDQ-Elternfragebogens (Strengths and Difficulties Questionnaire) für Screening und Diagnosestellung einer ADHS (Aufmerksamkeitsdefizit-Hyperaktivitätsstörung)

(published in: Kinderärztliche Praxis 77 (4) 230-235)

Andreas Becker, Veit Roessner, Aribert Rothenberger & Tobias Banaschewski

Abteilung für Kinder -und Jugendpsychiatrie/Psychotherapie, Universität Göttingen

Korrekturadresse: Andreas Becker
                            Universität Göttingen
                            Abteilung für Kinder -und Jugendpsychiatrie/
                            Psychotherapie
                            von-Siebold-Str. 5
                            D - 37075 Göttingen, Germany
                            Tel: ++49-551-392386
                            Fax: ++49-551-392696
                            Email: abecker4@gwdg.de
                            Webseite: http://www.gwdg.de/~ukyk

Wesentliches für die Praxis:

> ➢ Eine grundsätzliche kinderpsychiatrische Fragestellung ist die Einschätzung, ob überhaupt eine behandlungsbedürftige psychische Erkrankung vorliegt.
>
> ➢ Hierbei sollen ökonomische als auch valide Screeninginstrumente zur Diagnosestellung einer ADHS beitragen.
>
> ➢ Die Unterskala « Hyperaktivität » des Eltern SDQ wurde an einer Kliniksstichprobe auf ihre Eignung zur Diagnostik einer ADHS geprüft und stellte sich als sinnvolle Hilfe zur Vorhersage von entsprechenden Verhaltensauffälligkeiten dar.
>
> ➢ Es konnte gezeigt werden, dass bei einem unauffälligen Wert der Unterskala Hyperaktivität in 84% der Fälle keine Diagnose einer ADHS vorlag.
>
> ➢ Auffällige Werte konnten nur in 57% die Diagnose ADHS voraussagen.
>
> ➢ Die Vorhersage der Diagnose ADHS gelang bei jüngeren Kindern besser als bei älteren.

Hintergrund

Im Hinblick auf die Krankenversorgung von Patienten mit einer Aufmerksamkeitsdefizit-Hyperaktivitätsstörung (ADHS) im Praxisalltag von niedergelassenen Kinderärzten als auch Kinder- und Jugendpsychiatern sind Informationen von Seiten der Eltern durch Fragebögen eine wichtige Hilfe (2, 3).

Screening-Instrumente werden allerdings nicht zur individuellen Diagnosestellung eingesetzt, sondern dienen vielmehr als Vorauswahlverfahren, anhand dessen überprüft wird, ob überhaupt eine behandlungsbedürftige psychische Erkrankung vorliegt (5). Die daraus als auffällig hervorgegangenen Kinder und Jugendlichen werden anschließend einem Arzt zur weiteren differentialdiagnostischen Abklärung vorgestellt. Dabei stellt das ausführliche klinische Interview zusammen mit der Verhaltensbeobachtung die primäre Vorgehensweise zur Diagnosestellung der ADHS dar (9).

In verschiedenen Studien (7, 8) konnte aufgezeigt werden, dass hinsichtlich der Klassifikation kinderpsychiatrischer Störungen zwischen strukturierten Interviews und sorgfältig erarbeiteten Checklisten nur geringe Unterschiede bezüglich der Zuverlässigkeit und der Gültigkeit dieser Verfahren vorliegen.

Es erschien daher sowohl aus theoretischer als auch aus alltagspraktischer Sicht sinnvoll, die deutsche Version des Elternfragebogens zu Stärken und Schwächen eines Kindes (Strengths and Difficulties Questionnaire, SDQ) an einer klinischen Stichprobe (N=314) daraufhin zu prüfen, ob und mit welcher Güte der SDQ bzw. die Unterskala „Hyperaktivität" im Rahmen der Diagnostik einer ADHS zur Arbeitserleichterung und diagnostischen Sicherheit beitragen kann (3, 11).

Strengths and Difficulties Questionnaire (SDQ)

Der SDQ wurde 1997 in England als ein mehrdimensionales Screening-Instrument zur Erfassung kindlicher Stärken und Verhaltensauffälligkeiten bei 4 - 16 Jahre alten Kindern entwickelt (10). Er besteht aus 25 Items, die sich auf 5 Unterskalen verteilen. Vier Problemskalen behandeln emotionale Probleme, externalisierende Verhaltensauffälligkeiten, Hyperaktivität sowie Probleme mit Gleichaltrigen, während eine Skala prosoziales Verhalten abfragt.

Zusätzlich zu der Eltern- und Lehrerauskunft enthält dieser Fragebogen eine Version für die Kinder und Jugendlichen selbst (ab 11 Jahren). Im Gegensatz zu vorhergehenden Messinstrumenten (Rutter-Skala, CBCL-Child Behaviour Checklist) wurden die Items des SDQ sowohl nach Faktorenanalysen umfangreicher Frageninventare ausgewählt als auch an nosologischen Störungskonzepten orientiert. In Anlehnung an die Klassifikationssysteme DSM-IV (APA, 1994) (1) und ICD-10 (WHO, 1994) (12) wurden etwa für die Hyperaktivitäts Skala Items gewählt, die nach Überaktivität (2 Items), Aufmerksamkeitsstörungen (2 Items) und Impulsivität (1 Item) fragen.

Eltern, Lehrer und Jugendliche können den Fragebogen innerhalb von etwa 5 Minuten beantworten und Kinderärzte können mit Hilfe von Schablonen eine Auswertung innerhalb weniger Minuten leisten. Die vergleichende Betrachtung der drei Urteilsperspektiven ermöglicht einen zusätzlichen Informationsgewinn. In einer erweiterten Fragebogenversion werden zusätzlich globalere Schwierigkeiten in den Bereichen Stimmung, Konzentration und Verhalten im Umgang mit Anderen erfaßt, die die Aspekte Dauer, subjektiver Leidensdruck und einhergehende Beeinträchtigungen und Belastungen der Symptomatik differenzieren (10). Weiterhin sind für Mehrfachbefragungen entsprechende Follow-up-Versionen erhältlich.

Handhabung des SDQ

Auf dem Messinstrument wird jedes der 25 Items auf einer Drei-Punkte-Skala als entweder „teilweise zutreffend" mit „1" bewertet oder als „nicht zutreffend" bzw. „eindeutig zutreffend" je nach Merkmalspolung mit „0" bzw. „2" bewertet, so dass höhere Werte eine stärkere Ausprägung der durch die Skala gemessenen Eigenschaft anzeigen.

Für die Auswertung wird zuerst der Skalenwert der Unterskalen durch Addition der jeweils relevanten Items berechnet. Um den Skalenwert einer Unterskala zuverlässig bewerten zu können, müssen Angaben zu mindestens drei der fünf relevanten Merkmale vorhanden sein. Die Werte der gegebenen Antworten werden bei dem Fehlen von höchstens zwei Antworten aufsummiert, durch die Anzahl der vorhandenen Antworten dividiert und anschließend mit fünf multipliziert. Daraus ergeben sich dann Werte zwischen 0 und 10.

Anschließend wird der Gesamtproblemwert des SDQ berechnet, der zwischen 0 und 40 variieren kann. Er ergibt sich durch das Aufaddieren der Skalenwerte der vier Problemskalen. Die Skala "Prosoziales Verhalten" bleibt bei der Berechnung des Gesamtproblemwerts unberücksichtigt. Für den Gesamtproblemwert müssen somit mindestens 12 (mindestens 3 von 5 Antworten je Problem-Skala) der 20 relevanten Items beantwortet worden sein, um das Gesamtergebnis noch verwenden zu können. Zusätzlich werden auf der Rückseite des Fragebogens noch die subjektiven Belastungen des Kindes oder Jugendlichen erfasst.

Die gute Handhabbarkeit sind ebenso wie seine hohe Trennschärfe und die mehrfach belegte psychometrische Güte (5, 10, 13) dazu geeignet, den SDQ als ein hilfreiches Screening-Instrument im Rahmen der Diagnostik für ADHS im klinischen Alltag einzusetzen (3, 6).

Der SDQ ist in etwa 40 Sprachen verfügbar und kann über die Internetadresse www.sdqinfo.com heruntergeladen und für nicht-kommerzielle Zwecke kostenfrei benutzt werden. Auswertungsanleitungen und - schablonen, Normen sowie eine Literaturliste zum SDQ sind unter http://wwwuser.gwdg.de/~ukyk/inventare.html zu finden.

Anwendungsbereiche und Güte

Der SDQ wird sowohl als Screening-Instrument für epidemiologische Studien als auch zur klinischen Diagnostik, zur Evaluierung von Therapieverläufen und für wissenschaftliche Fragestellungen verwendet. Die Normierung und Evaluation der deutschen SDQ -Elternversion wurde zwischen März 1998 und Februar 1999 von Woerner et al. (13, 14) an einer repräsentativen bundesweiten Stichprobe durchgeführt und 2002 veröffentlicht.

Faktorenstruktur

Im Rahmen einer Studie (5) zur Überprüfung der Güte der psychometrischen Eigenschaften zeigte sich, dass mittels einer Faktorenanalyse die vorgegebenen SDQ-Skalen exakt repliziert werden konnten, d.h., dass alle Items einer jeden SDQ-Skala jeweils gemeinsam auf einem Faktor luden und außerdem sämtliche Ladungen exakt die intendierte Skalenzugehörigkeit widerspiegelten.

Reliabilität (Cronbach's Alpha)

Die Zuverlässigkeit der Messung des „Gesamtproblemwerts" (Cronbach's α) beträgt .82 und kann somit als gut angesehen werden; die Homogenitätswerte der einzelnen Unter-Skalen erreichten dabei Ergebnisse zwischen .58 und .76. (5).

Rohwertverteilung

Entsprechend der Vorgehensweise beim englischen SDQ wurden Grenzwerte ermittelt, mit denen die Rohwerte jeweils zu einer von drei Kategorien (unauffällig / grenzwertig / auffällig) zuzuordnen waren. Wie bei der englischen SDQ-Grenzwertbestimmung wurde dabei von einem Anteil von 10% auffälligen und 80% unauffälligen Werten ausgegangen, die von einer dazwischen liegenden Kategorie mit 10% Grenzbefunden getrennt wurden.

Bei einer zusätzlichen Analyse der Alters- und Geschlechtsunterschiede ließen sich nur geringfügige Abweichungen finden, so dass eine alters- und / oder geschlechtsspezifische Auswertung nicht erforderlich ist. Die deutschen Normwerte sind in Tabelle 1 dargestellt.

Tabelle 1: Cut-Off-Werte, Eltern SDQ (Deutsche Normen; N:930; 6-16 Jahre)

SDQ Skala	„unauffällig"	„grenzwertig"	„auffällig"
Gesamtproblemwert	0 - 12	13 - 15	16 - 40
Emotionale Probleme	0 - 3	4	5 - 10
Verhaltensprobleme	0 - 3	4	5 - 10
Hyperaktivität	0 - 5	6	7 - 10
Probleme mit Gleichaltrigen	0 - 3	4	5 - 10
Prosoziales Verhalten	6 - 10	5	0 - 4

Die Werte stellen Rohwerte der jeweiligen SDQ- Skala dar

Sensitivität, Spezifität und positiver und negativer prädiktiver Wert

Die für den Kliniker wichtige Frage lautet, wie viele der durch den Eltern SDQ als grenzwertig bzw. auffällig beschriebenen Kinder auch tatsächlich eine entsprechende Diagnose (hier: ADHS) aufweisen (positiver prädiktiver Wert) bzw. wie viele der als unauffällig beschriebenen Kinder keine entsprechende Diagnose erhalten haben (negativer prädiktiver Wert). Diese Testparameter werden ganz wesentlich auch durch die Häufigkeit der Störung in der jeweils untersuchten Patientengruppe bestimmt: Ist die Erkrankung sehr selten wird im Vergleich zu einer häufigeren Erkrankung bei ansonsten konstanten Testeigenschaften der positive prädiktive Wert erniedrigt und der negative prädiktive Wert erhöht und umgekehrt. Weiterhin ist von Bedeutung, und dies sind die eigentlichen Testmerkmale, wie hoch der Anteil der von einem Arzt / Facharzt diagnostizierten Kinder mit ADHS ist, die gleichzeitig einen grenzwertigen bzw. auffälligen Skalenwert auf der Unterskala „Hyperaktivität"

erhalten haben (Sensitivität), bzw. wie viele der Kinder die keine ADHS Diagnose erhielten, einen entsprechend unauffälligen Skalenwert aufwiesen (Spezifität).

Eigene Untersuchungen

Zur Beantwortung dieser Fragen wurde eine Teilstichprobe der kinderpsychiatrischen Inanspruchnahmepopulation der Göttinger Universitätspoliklinik für Kinder- und Jugendpsychiatrie/Psychotherapie zur Überprüfung der Anwendungsgüte der Elternversion des SDQ untersucht. Nach ausführlicher klinischer Untersuchung durch Fachärzte für Kinder- und Jugendpsychiatrie/-psychotherapie wurde bei 126 von 314 Kindern (normal intelligent; Alter: 5 – 13 Jahre) die Diagnose einer ADHS gestellt, während bei 188 Kindern keine ADHS, aber andere kinder- und jugendpsychiatrische Probleme vorlagen. Somit zeigte sich für diese kinderpsychiatrische Inanspruchnahmestichprobe eine Prävalenz der ADHS von ca. 40%. Für diese Stichprobe wurden der positive und der negative prädiktive Wert sowie die Sensitivität und die Spezifität für den in der Normierung ermittelten Cut-Off Wert für die Skala „Hyperaktivität" berechnet. Um eventuelle Alterseffekte zu berücksichtigen, wurde die Gesamtgruppe in drei Altersgruppen unterteilt.

Der positive prädiktive Wert (**PPW**) eines Tests beschreibt die Wahrscheinlichkeit, mit der ein Kind oder ein Jugendlicher mit einem positiven Testergebnis auch tatsächlich „krank" ist (**rp / (rp + fp)**) und der negative prädiktive Wert (**NPW**) ist die Wahrscheinlichkeit mit der ein Patient mit einem negativen Testbefund auch tatsächlich „gesund" ist (**rn / (rn + fn)**); siehe dazu Tabelle 2. Diese Prädiktionswerte sind im Gegensatz zu den Parametern Sensitivät und Spezifität stark abhängig von

der Häufigkeit der Prävalenz der Zielgruppe in der jeweiligen Untersuchungspopulation.

Tabelle 2: Vierfeldertafel mit positiven & negativen prädiktiven Wert sowie Sensitivität & Spezifität

<div align="center">Zustand des Patienten</div>

	krank	**nicht krank**	
Testergebnis positiv	Richtig positiv **(rp)**	Falsch positiv **(fp)**	**„positiver prädiktiver Wert"** **rp / (rp + fp**)
Testergebnis negativ	Falsch negativ **(fn)**	Richtig negativ **(rn)**	**„negativer prädiktiver Wert"** **rn / (rn + fn**)
	„Sensitivität" **rp** / (**rp + fn**)	**„Spezifität"** **rn / (rn + fp**)	

Die Sensitivität ist der Anteil der richtig positiv erkannten Kinder und Jugendlichen von allen „Kranken" (**rp** / (**rp + fn**)). Die Spezifität beschreibt den Anteil der richtig negativ erkannten Kinder und Jugendlichen an den „Nicht-Kranken" (**rn / (rn + fp**)). Sowohl die Sensitivität als auch die Spezifität sind weitgehend unabhängig von der Häufigkeit des Vorkommens der Störung in der jeweiligen Untersuchungsgruppe.

Für die Berechnung dieser Kenngrößen wurden - aus Gründen der Übersichtlichkeit - die Kinder und Jugendlichen mit einem „grenzwertigen" Urteil der Kategorie „auffällig" zugeordnet.

Aus der Tabelle 3 sind altersspezifische positive und negative prädiktive Werte sowie die altersspezifische Sensitivität und Spezifität für die SDQ Skala „Hyperaktivität" zu entnehmen. Es zeigte sich für die Gesamtgruppe, dass ein „unauffälliger Wert" auf der SDQ Skala „Hyperaktivität" in 84% (NPW) der Fälle mit der Abwesenheit der entsprechenden klinischen Diagnose ADHS (richtige negative Zuordnung) verbunden ist. Allerdings hat nur bei 57% der Kinder mit auffälligen Werten auf der Skala „Hyperaktivität" ein ADHS vorgelegen (PPW).

Tabelle 3 : Sensitivität, Spezifität, positiver und negativer prädiktiver
Wert der SDQ Skala „Hyperaktivität"

(ADHS ja / nein)	Sens.	Spez.	PPW	NPW
Alle Patienten: 5 - 13 Jahre (N = 126/188)	84 %	56 %	57 %	84 %
Altersgruppe: 5 - 7 Jahre (N = 34/33)	91 %	64 %	72 %	88 %
Altersgruppe: 8 - 10 Jahre (N = 54/89)	85 %	61 %	57 %	87 %
Altersgruppe: 11 - 13 Jahre (N = 38/66)	76 %	49 %	46 %	78 %

Sens: Sensitivität; Spez: Spezifität; PPW: positiver prädiktiver Wert; NPW: negativer prädiktiver Wert

Die Wahrscheinlichkeit einer richtigen negativen Zuordnung lag in allen untersuchten Altersstufen ähnlich hoch, auch wenn dabei die Gruppe der 11 - 13 jährigen etwas schlechter abschnitt (88%, 87%, 78%). Dem gegenüber fielen die Prozentwerte für eine richtige positive Zuordnung aufgrund auffälliger Werte der Unterskala „Hyperaktivität" zur klinischen Diagnose ADHS mit dem Älterwerden stetig geringer aus (72%, 57%, 46%).

Ergänzend wurden die Sensitivität und Spezifität für die Elternberichte bestimmt.
Dadurch wurde aufgezeigt, wie viele derjenigen Kinder, die von Therapeutenseite als
„Fälle" klassifiziert worden sind, ebenfalls einen auffälligen Wert auf der SDQ Skala
„Hyperaktivität" erhalten haben (Sensitivität) und wie viele Kinder und Jugendliche,
die keine entsprechende Diagnose durch den Therapeuten erhielten, auch durch die
Fragebögen als unauffällig beurteilt werden (Spezifität).

Der Anteil der richtig als „auffällig" erkannten Kinder und Jugendlichen (Sensitivität)
lag für die gesamte Stichprobe für die Skala „Hyperaktivität" bei 84%. Am besten war
die richtige Zuordnung für die Gruppe der 5-7-jährigen Kinder mit 91%. Etwas
niedrigere Werte fanden sich für die beiden anderen Altersgruppen (8-10 Jahre: 85%
und 11-13 Jahre 76%). Im Gegensatz dazu fiel der Anteil der richtig als „unauffällig"
erkannten Kinder und Jugendliche (Spezifität) für die gesamte Stichprobe mit 56%
deutlich niedriger aus.

Schlußfolgerung

In der vorliegenden Untersuchung wurde ein Elternfragebogen (SDQ) bei 314
Jungen einer kinderpsychiatrischen Inanspruchnahmestichprobe (126 mit ADHS/188
ohne ADHS) eingesetzt, um zu prüfen, ob die Unterskala „Hyperaktivität" des SDQ
bei der klinischen Abklärung einer ADHS im Praxisalltag hilfreich sein und zur
geforderten diagnostischen Qualitätsverbesserung beitragen kann.

Es konnte gezeigt werden, dass eine erhebliche Anzahl der Kinder und
Jugendlichen, die eine Diagnose ADHS erhielten, ebenso durch die SDQ-Unterskala

„Hyperaktivität" als auffällig beurteilt worden sind. Diese Leistung der Unterskala „Hyperaktivität" nahm von den jüngeren (5 - 7 Jahre) zu den älteren (11 - 13 Jahre) Kindern nur geringfügig ab (Sensitivität: 91% vs. 76%). Von den Kindern, die keine Diagnose erhielten, wurde jedoch nur in 56% eine Übereinstimmung mit der Beurteilung durch den SDQ erzielt (niedrige Spezifität). Daraus ist zu schließen, dass nur etwa die Hälfte der Kinder und Jugendlichen, die keine relevante Störung im Sinne einer ADHS haben, auch im Elternurteil als diesbezüglich „unauffällig" klassifiziert werden.

Insgesamt zeigen die Ergebnisse, dass der SDQ-Elternfragebogen hinsichtlich einer positiven Zuordnung zur Diagnose ADHS eher bei jüngeren Kindern eine Hilfe im Praxisalltag darstellt als bei älteren Kindern. Es wäre also insbesondere bei den älteren Kindern zu prüfen, ob zusätzliche von ihnen selbst und/oder den Lehrern ausgefüllte Fragebögen (z.B. SDQ für Lehrer und Kinder ab 11 Jahre) die diskriminante Validität erhöhen können.

Die Ergebnisse machen deutlich, dass bei unauffälligen Werten auf der SDQ Skala „Hyperaktivität" eine sehr gute Voraussage in dem Sinne möglich ist, dass bei solchen Kindern wahrscheinlich kein ADHS vorliegt (84%). Dies gilt offenbar gleichermaßen im gesamten Altersbereich von 5 - 13 Jahren (78 - 88% richtig negative Zuordnung). Demgegenüber sind auffällige Werte der Unterskala „Hyperaktivität" im SDQ nur bedingt eine Hilfe, um daraus eine ADHS im Sinne einer richtig positiven Zuordnung vorherzusagen (57%). Weiterhin zeigte sich ein deutlicher Alterseffekt bei der Vorhersage einer ADHS-Diagnose: Während bei den 5

- 7 jährigen Kindern mit 72% noch ein recht guter Wert vorlag, sank die Wahrscheinlichkeit bei den 8 - 10jährigen auf 57% und bei den 11 - 13jährigen sogar auf 46%. Daher muss man mit einem solch auffälligen Befund der Unterskala „Hyperaktivität" im Eltern-SDQ sehr zurückhaltend umgehen und darf keinesfalls voreilig die Diagnose ADHS stellen. Aus diesem Grund ist es wichtig, dem Screening-Verfahren eine weiterführende umfangreiche klinische Abklärung entsprechend der diagnostischen Leitlinien für ADHS (9) folgen zu lassen.

Insgesamt kann gesagt werden, dass die Unterskala „Hyperaktivität" des Eltern-SDQ mit seinen ökonomischen und dennoch fachlich gut fundierten Eigenschaften eine sinnvolle Hilfe darstellt, um im Rahmen eines Screenings von Verhaltensauffälligkeiten zur Verbesserung der Diagnostik im Praxisalltag beizutragen. Damit haben wir mit dem SDQ ein Untersuchungsinstrument an der Hand, welches die empirisch geleitete evidenz-basierte Eingangs- und Verlaufsdiagnostik im Hinblick auf Verhaltensprobleme bereichern, jedoch eine störungsspezifische Anamnese und klinische Untersuchung mit Verhaltensbeobachtung nicht ersetzen kann.

Beim Einsatz in der Kinderarztpraxis muss bei der Beurteilung der Testergebnisse berücksichtigt werden, bei welchen Patienten der Test eingesetzt werden soll. Der Anteil der durch einen Test als richtig erkannten, tatsächlich Kranken hängt nicht nur von der Güte des Tests ab, sondern ganz wesentlich von der Ausgangs-wahrscheinlichkeit für die Erkrankung in der untersuchten Stichprobe. Nehmen wir folgendes fiktives Beispiel (persönliche Mitteilung Prof. Kries):

Die Ausgangswahrscheinlichkeit für eine HIV-Infektion bei Nonnen sei geschätzt 0.0001%, also annähernd null, während diese bei Prostituierten bei etwa 40% liege. Setzen Sie in beiden Gruppen einen HIV Test – mit sehr hoher Sensitivität (99%) und Spezifität (99%) ein, so wird die Rate der als richtig HIV- positiv vorhergesagten Nonnen sehr gering ausfallen. Obwohl auch bei den Nonnen einige Tests positiv sein werden - immerhin ist der Test nur ca. 99% spezifisch - können diese weiterhin ruhig schlafen, - nur eine von 500 000 positiv getesteten Nonnen wird tatsächlich HIV-infiziert sein. Bei ihnen besagt das positive Testergebnis also fast nichts. Für die Prostituierten hingegen ergibt sich eine Infektionswahrscheinlichkeit von über 99.84% (PPW) für die durch den Test als HIV-positiv vorhergesagten Frauen. Obwohl genau der gleiche Test verwendet wurde, ist in diesem Fall davon auszugehen, dass jede testpositive Prostituierte auch tatsächlich HIV infiziert ist.

Was bedeutet dies für den Einsatz des SDQ in der Kinderarztpraxis? Setzen Sie den Test als Screening bei allen Kindern ein, ist die a priori Wahrscheinlichkeit, dass diese Kinder ein ADHS haben, so hoch wie in der Allgemeinbevölkerung, nämlich ca. 5%. Setzen Sie den Test jedoch bei Kindern ein, bei denen Sie, die Eltern oder die Erzieherin den Verdacht auf ein ADHS haben, wird allein aufgrund dieses begründeten Verdachts die a priori Wahrscheinlichkeit - d.h. die Rate der tatsächlich erkrankten Kinder vor der Testdurchführung - schon annähernd so hoch sein (ca. 30%), wie bei den in der Kinder- und Jugendpsychiatrie in Göttingen untersuchten Kindern.

Hieraus ergibt sich:

> ➤ Wird der SDQ als Screening Test - also bei allen Kindern - eingesetzt, ist davon auszugehen, dass bei bis zu 5% der Kinder ein ADHS vorliegt. In diesem Fall erhöht ein positives Testergebnis die Wahrscheinlichkeit, dass ein

ADHS vorliegt erheblich: etwa jedes 10. im Test auffällige Kind ist tatsächlich betroffen.

➢ Setzen Sie den Test bei einem Kind ein, bei dem Sie, die Eltern oder Erzieherin den Verdacht auf ADHS haben ist die „a priori Wahrscheinlichkeit" für ADHS deutlich höher. Wie hoch, ist im Einzelfall schwer vorauszusagen. Grob geschätzt wird sie ähnlich dem in der in Göttingen in der Kinder- und Jugendpsychiatrie untersuchten Patientenkollektiv sein. Also können Sie einen positiven prädiktiven Wert von 46% - 72%, je nach Alter, bzw. negativen prädiktiven Wert von 78% -88% erwarten: 46% - 72% der testpositiven Kinder werden tatsächlich ein ADHS haben und 78% -88% der testnegativen Kinder werden kein ADHS haben, obwohl Sie, die Eltern oder Erzieherin den Verdacht auf ADHS hatten.

➢ Mit der gegebenen Hintergrundinformationen kann der SDQ die diagnostische Arbeit, im Hinblick auf psychische Störungen, in der Kinderarztpraxis zielgerichtet führen und erleichtern.

Textkasten:

Relevanz von SDQ Ergebnissen in der Praxis

➢ Besonders bei Kindern im Vorschulalter werden die meisten tatsächlich an ADHS erkrankten Kinder durch den Test identifiziert.

➢ Wird der Test als Screening für alle Kinder eingesetzt

   o Sind bei auffälligem Befund nur etwa 10% der betreffenden Kinder erkrankt

➢ Wird der Test bei Kindern mit wie auch immer begründetem Verdacht auf ADHS eingesetzt, kann nur eine grobe Vorhersage über den Anteil der tatsächlichen Fälle unter den testpositiven und der nicht ADHS Fälle unter den testnegativen gemacht werden.

- o Sind bei auffälligem Befund ca. 60% oder mehr der betreffenden Kinder auch erkrankt

- o Haben bei unauffälligem Befund mindestens 80% der betreffenden Kinder kein ADHS

LITERATUR

1. American Psychiatric Association DSM IV (1994). Diagnostic and Statistical. Manual of Mental Disorders. 4th ed. Washington, DC

2. Arbeitsgruppe Deutsche Child Behavior Checklist (1998). Elternfragebogen über das Verhalten von Kindern und Jugendlichen: Deutsche Bearbeitung der Child Behavior Checklist (CBCL/4-18). Einführung und Anleitung zur Handauswertung. 2. Auflage mit deutschen Normen. Arbeitsgruppe Kinder-, Jugend- und Familiendiagnostik. Köln

3. Banaschewski T, Woerner W, Becker A, Rothenberger A (2004). Diagnostik der Aufmerksamkeitsdefizit-Hyperaktivitäts-Störung: Unterstützung durch den Elternfragebogen zu Stärken und Schwächen des Kindes (SDQ). Monatsschrift Kinderheilkunde 152 (7): 778-781

4. Barkley et al. (2002). Consensus Statement on ADHD. Eur Child Adolesc Psychiatry 11: 96-98

5. Becker A, Woerner W, Hasselhorn M, Banaschewski T, Rothenberger A (2004). Validation of the parent and teacher SDQ in a clinical sample. Eur Child Adolesc Psychiatry 13 (2): 11-16

6. Becker A, Hagenberg N, Roessner V, Woerner W, Rothenberger A (2004). Evaluation of the self-reported SDQ in a clinical setting: Do self-reports tell us more than ratings by adult informants? Eur Child and Adolesc Psychiatry 13 (2) 17-23

7. Boyle MH, Offord DR, Racine Y, Szatmari P, Fleming JE; Sanford M (1996). Identifying thresholds for classifying childhood psychiatric disorders: issues and prospects. J Am Acad Child Adolesc Psychiatry 35: 1440-1448

8. Boyle MH, Offord DR, Racine Y, Szatmari P, Sanford M, Fleming JE (1997). Adequacy of interviews vs checklists for classifying childhood psychiatric disorders based on parent reports. Arch Gen Psychiatry 54: 793-799

9. Deutsche Gesellschaft für Kinder- und Jugendpsychiatrie und Psychotherapie (Hrsg.) (2000) Leitlinien zur Diagnostik und Therapie von psychischen Störungen. Deutscher Ärzte-Verlag Köln. www.uni-duesseldorf.de/WWW/AWMF

10. Goodman R (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. J Am Acad Child Adolesc Psychiatry 40: 1337-1345

11. Klasen H, Woerner W, Rothenberger A, Goodman R (2003). Die deutsche Fassung des Strengths and Difficulties Questionnaire (SDQ-Deu) - Übersicht und Bewertung erster Validierungs- und Normierungsbefunde. Prax Kinderpsychol Kinderpsychiatr 7: 491-502

12. WHO - World Health Organization (1992). The ICD-10 classification of mental and behavioural disorders. Clinical Descriptions and diagnostic guidelines. WHO, Geneva

13. Woerner W, Becker A, Friedrich C, Klasen H, Goodman R, Rothenberger A (2002). Normierung und Evaluation der deutschen Elternversion des Strengths and Difficulties Questionnaire (SDQ): Ergebnisse einer repräsentativen Felderhebung. Z Kinder Jugendpsychiatr Psychother 30: 105-112

14. Woerner W, Becker A, Rothenberger A (2004). Normative data and scale properties of the German parent SDQ. Eur Child and Adolesc Psychiatry 13 (2) 3-9

4.4 Psychopathological screening of children with ADHD: Strengths and Difficulties Questionnaire in a pan-European study

(published in: European Journal of Child and Adolescent Psychiatry 15 (1) 56-62)

Authors: Andreas Becker[1], Hans-Christoph Steinhausen[2], Gisli Baldursson[3], Sören Dalsgaard[4], Maria J Lorenzo[5], Stephen J Ralston[6], Aribert Rothenberger[1✉] and the ADORE study group*

Affiliations:

[1]Department of Child and Adolescent Psychiatry, University of Göttingen, Germany

[2]Department of Child and Adolescent Psychiatry, University of Zürich, Switzerland

[3]Department of Child and Adolescent Psychiatry, University of Reykjavík, Iceland

[4]Department of Child and Adolescent Psychiatry, University of Aarhus, Denmark

[5]Eli Lilly and Company, Lilly Research Centre, Windlesham, Surrey, UK

[6]Employed by Eli Lilly & Co. at the time the research was performed

Members of the ADORE study group:

G Baldursson, D Coghill, P Curatolo, S Dalsgaard, M Döpfner, B Falissard, A Hervas, MF Le Heuzey, TS Nøvik, RR Pereira, U Preuss, S Ralston, P Rasmussen, AW Riley, A Rothenberger, G Spiel, HC Steinhausen, L Vlasveld

Corresponding Author:

Prof. Dr. Aribert Rothenberger

University of Göttingen

Child and Adolescent Psychiatry

von-Siebold-Str. 5

D-37075 Göttingen, Germany

Phone: +49-551-396727

Fax: +49-551-398120

email: arothen@gwdg.de

website: http://www.gwdg.de/~ukyk

Abbreviations

ADHD	Attention-Deficit/Hyperactivity Disorder
ADORE	Attention-deficit/hyperactivity Disorder Observational Research in Europe
ADHD-RS-IV	ADHD Rating Scale-IV
CHIP-CE	Child Health and Illness Profile - Child Edition
CGAS	Children's Global Assessment Scale
CGI-S	Clinical Global Impression - Severity scale
SDQ	Strengths and Difficulties Questionnaire

Abstract

Objective: To examine the psychometric properties of the Strengths and Difficulties Questionnaire (SDQ) parent version and to determine the effects of age, gender, country and investigator type (paediatrician, child psychiatrist, other physician) on the SDQ scores in the prospective, non-interventional ADORE study.

Methods: The SDQ was completed for 1,459 children with ADHD (aged 6–18 years) in 10 European countries.

Results: Factor analysis provided an exact replication of the original 5-factor SDQ subscale structure. All subscales were sufficiently homogeneous. The mean total difficulties and SDQ subscale scores of the ADORE sample clearly differed from UK normative data. Younger children were more impaired on different SDQ scales than older children, and girls were more emotionally affected than boys. Differences between countries were found for each SDQ scale, but the investigator type had no significant effect. Correlation coefficients between SDQ scales and other scales used in ADORE ranged from low ($r<0.30$) to high ($r>0.50$).

Conclusions: The present study confirmed the validity and reliability of the parent-reported SDQ scale structure and showed that the scale scores are dependent on age and gender. In contrast to investigator type, different cultures had a significant effect on SDQ scores. Correlations with other scales used in the ADORE study underline both separate domains and meaningful associations.

Introduction

Children with attention-deficit/hyperactivity disorder (ADHD) suffer not only from the core symptoms of the disorder, but also from several co-existing psychiatric problems. However, it is only recently that "co-morbidity" in ADHD has come to the

forefront as one of the most important aspects of the disorder. As many as two-thirds of all individuals with ADHD in the general population meet the criteria of at least two additional diagnoses, e.g. oppositional defiant disorder, conduct disorder or depression and anxiety disorder, which may influence peer relationships, social behaviour, psychosocial functioning, and health-related quality of life of children with ADHD .

Assessment of children with ADHD, therefore, should attempt to cover a broad range of different behavioural aspects beyond the core symptoms to establish a complete psychopathological profile as a prerequisite for careful planning of a multimodal treatment programme. In this respect, the European guidelines for hyperkinetic disorder recommend the Strengths and Difficulties Questionnaire (SDQ) as a useful assessment instrument supplementary to clinical interview .

The main clinical advantages of the SDQ are: its briefness and comprehensiveness; the simple administration and scoring of the well-balanced number of negative and positive behavioural attributes it addresses; the correspondence of its subscales and items to the major categories and criteria of current psychiatric classification systems; the use of the same items and scales in the parent, teacher and self-report versions; and its ready availability (in more than 40 languages and free of charge for non-commercial purposes) by downloading from the internet (www.sdqinfo.com). On the other hand, several mono-symptomatic disorders (e.g. tic disorders) and a range of more specific child psychiatric problems (including eating disorders or enuresis) are not explicitly covered by the 25 core items of the SDQ .

The psychometric reliability and validity of the SDQ subscales for each language and country needs to be examined, in addition to establishing representative national norms. Since the latter might differ from country to country, an international study like

ADORE (Attention-deficit/hyperactivity Disorder Observational Research in Europe) needs to evaluate the psychometric properties of the SDQ before performing detailed analyses on the main hypotheses of this naturalistic study on ADHD treatment in children.

The aim of the present paper is to examine the psychometric properties of the SDQ within the framework of the ADORE study, and to evaluate potential differences with regard to age, gender, country, and investigator type (paediatricians, child psychiatrists, other physicians). Although the SDQ was completed by parents, setting effects cannot be ruled out . Furthermore, we examined the correlations between the SDQ subscales and other instruments/questionnaires used in the ADORE study.

Methods

Procedure

The ADORE study is a prospective, non-interventional study of approximately 1,500 patients observed by approximately 300 investigators in 10 European countries. The primary objective of ADORE is to describe the relationship between treatment regimen prescribed and quality of life in children with ADHD over a 2-year period. The naturalistic care provided and the outcomes (psychopathology, quality of life) are recorded at seven data collection points . The present report provides data of the SDQ on all patients at the first data collection point (baseline) before any treatment had been initiated for ADHD symptoms.

Sample

A total of 1,573 children with hyperactive/inattentive/impulsive symptoms but no previous formal diagnosis of ADHD and no mental retardation, autism or schizophrenia were enrolled in the ADORE study. Of the 1,478 patients eligible for the baseline analysis, complete SDQ data sets were available for 1,459 children. Gender information was available for 1,453 children (231 girls and 1,222 boys). Mean age of the girls was 8.8 years (standard deviation, SD 2.3) and mean age of the boys was 9.0 years (SD 2.5).

Instruments

The SDQ parent-reported version was used. This short behavioural screening questionnaire consists of 25 items and can be completed within about five minutes by parents of children aged 4 to 16 years. The SDQ contains five different subscales (of 5 items each) measuring: emotional symptoms, conduct problems, hyperactivity-inattention, peer relationship problems and prosocial behaviour. Each of the 25 items of the SDQ is scored on a 3-point scale with 0= not true, 1=somewhat true or 2=certainly true, with a higher score indicating greater problems except for prosocial behaviour, where a higher score indicates more positive behaviour. A total difficulties score (range 0–40) is obtained by summing the scores of the emotional symptoms, conduct problems, hyperactivity-inattention and peer problems subscales.

The SDQ has been standardised in several countries (here we refer to the large representative sample of the United Kingdom, UK). The psychometric properties of the SDQ are satisfactory and its subscales are valid with corresponding child psychiatric diagnostic categories .

More detailed information on the additional scales used both in ADORE and in the present analysis, i.e. the ADHD-Rating Scale-IV (ADHD-RS-IV), the Child Health and Illness Profile-Child Edition (CHIP-CE), the Clinical Global Impression-Severity (CGI-S) scale and the Children's Global Assessment Scale (CGAS), is provided in other articles of this issue .

Statistical analyses

A principal component factor analysis with varimax rotation was performed to test the validity of the internal structure of the SDQ. To allow for a direct comparison with the proposed factorial structure, the number of extracted factors was fixed *a priori* at five. Cronbach's alpha (α) coefficients were computed to estimate the internal consistency of the various SDQ subscales.

Descriptive statistics (mean and SD) for the SDQ subscales and the total difficulties scores are presented for the total sample and by age and gender. The mean (SD) scores for the SDQ subscales and total difficulties score were also calculated for each country participating in the study.

Analysis of variance (ANOVA) modelling was performed to determine the effects of age, gender, country and investigator type on the parent-rated SDQ subscale scores and total difficulties score. In addition, based on the UK norms (available at www.sdqinfo.com), the proportion of girls and boys with parent-rated SDQ scores categorised as 'normal', 'borderline' or 'abnormal' was calculated. Abnormal scores were defined as: total difficulties ≥ 17; conduct problems ≥ 4; hyperactivity-inattention ≥ 7; emotional symptoms ≥ 5; peer problems ≥ 4; and prosocial behaviour ≤ 4.

Finally, correlation coefficients between the parent-reported SDQ subscales and total difficulties score and the other scales in the study (ADHS-RS-IV, CHIP-CE, CGI-S, and CGAS) were calculated.

Results

Factorial validity and reliability

After rotation of the five extracted factors with initial eigenvalues >1, the resulting pattern of main loadings was an identical replication of the original SDQ subscales. The 5-factor solution explained 47% of the total variance. Table 1 shows the high correspondence between the pattern of rotated loadings and the original SDQ subscales. All 25 items had their highest loadings on the extracted factors and matched identically the intended scale structure. For the entire European sample, all factors were well-defined and no items were misclassified into different subscales. Factor analysis based on the country subsamples showed similar results (data not shown). However, some misclassifications occurred due to the small number of patients in some of the countries.

Internal consistency coefficients (Cronbach's α) for the SDQ subscales are also reported in Table 1. Although each subscale consists of only 5 items, the Cronbach's alpha coefficients were quite high (α=0.58–0.72) in the evaluated sample. This was fairly consistent for all countries (data not shown). The lowest value was found for the subscale measuring hyperactivity-inattention, and the highest alpha coefficients were found for the prosocial behaviour and peer problems subscales. Thus, the data of the parent-rated version of the SDQ in this sample can be considered sufficiently reliable.

[insert Table 1 here]

Moderating effects

Mean SDQ subscale scores and the total difficulties score for the ADORE sample clearly differed from the representative UK normative data as shown in Table 2. Age and gender effects are also shown in this table, whereas the mean SDQ subscale scores and total difficulties scores by country are summarised in Table 3. Data for "investigator type" were not significant and are not reported further.

The analysis of the age effects showed that younger children (aged 6–10 years) had significantly higher total difficulties, hyperactivity-inattention and peer relationship problems scores than older children (aged 11–18 years). Furthermore, the analysis of the gender effects showed that girls had significantly more emotional symptoms than boys, and exhibited significantly more prosocial behaviour than boys. Table 2 also presents the percentages of patients (girls and boys) who could be classified as abnormal (based on the UK cut-off scores) on each SDQ subscale and the total difficulties score.

[Insert Tables 2 & 3 here]

Country (df=9; F=13.69; p≤0.001) had a significant effect on the SDQ total difficulties score. Adjusted means for each category in the model were calculated but are not presented here. Comparisons of the descriptive statistics between the different European countries showed that children with ADHD from France and the UK had the highest scores on the SDQ total difficulties scale, while children from Iceland and Switzerland demonstrated the lowest scores (Table 3).

Country had a statistically significant influence (df=9; F=3.75; p≤0.001) on the emotional symptoms score. Children with ADHD in Denmark and France presented

with the highest scores on this subscale, whereas children from Italy and Norway had lower levels of emotional symptoms (Table 3).

When modelling the SDQ conduct problems subscale, country had a statistically significant effect (df=9; F=15.78; p≤0.001). Once again, children from France and the UK were the most strongly affected by conduct problems; the least affected children came from Norway and Iceland (Table 3).

Country (df=9; F=17.62; p≤0.001) had a significant effect on the SDQ hyperactivity-inattention subscale scores. Again, children from France, Denmark and the UK were the most impaired, and children from Iceland and Switzerland were the least affected (Table 3).

The results for the SDQ peer relationship problems subscale were similar to those found for the hyperactivity-inattention subscale score. Country had a significant effect (df=9; F=5.35; p≤0.001) on the SDQ peer relationship problems score. Again, children from France, Denmark and the UK were the most affected, whereas children from Iceland and Switzerland had the least peer relationship problems (Table 3).

Country (df=9; F=5.70; p≤0.001) had a significant effect on the SDQ prosocial behaviour subscale score. The lowest scores were seen in children from the UK (Table 3).

In summary, many of the SDQ subscale scores and the total difficulties score are affected by different moderating factors. Despite being statistically significant, the age and gender effects were quite small and may not be clinically meaningful. Statistically significant differences between countries were found for each SDQ subscale score and the total difficulties score.

Correlations

Table 4 shows the *pairwise correlations* between the SDQ subscale scores and the scores obtained with other scales used in the ADORE study. Many of the correlation coefficients were quite small (<0.3), indicating little overlap between the scales. On the other hand, the correlation coefficients for those scales that are closely related to ADHD were higher. The highest inverse correlations were found between the SDQ conduct problems subscale and the Risk Avoidance domain of the CHIP-CE scale (r = -0.73). There were moderate positive correlations between the SDQ hyperactivity-inattention subscale and the total score of the ADHD-RS-IV (r = 0.51) and the ADHD-RS-IV hyperactivity-impulsivity subscale score (r = 0.54). Similarly, there were moderate correlations between the SDQ conduct problems subscale score and the ADHD-RS-IV total score and hyperactivity-impulsivity subscale scores (both r = 0.42). The negative correlation of the SDQ conduct problems subscale score with the CHIP-CE Achievement (r = -0.41) domain score is also noteworthy, as are the correlations between the SDQ emotional symptoms subscale and CHIP-CE Satisfaction (r = -0.50) and Comfort (r = -0.57) domains. Finally, prosocial behaviour was positively correlated to CHIP-CE Resilience and Risk Avoidance domains (r = 0.41 and r = 0.40, respectively), while peer problems correlated negatively with the CHIP-CE Achievement domain (r = -0.58). Due to large sample size, all correlations between the SDQ scales and other scales were highly significant (p<0.001).

[Insert Table 4 here]

Discussion

Because ADHD is associated with various behavioural problems, knowledge of the entire psychopathological profile is important for further evaluation and treatment. In this respect, the SDQ is a useful and easy applicable questionnaire for screening purposes [9,10]. Hence, in the present naturalistic study [4, 6], the SDQ was obtained in 1,459 children with ADHD to obtain additional psychopathological information and to test for the psychometric properties of the SDQ.

Firstly, factor analysis was performed to test the validity of the SDQ. The factor structure of the 25 items of the SDQ representing the original five subscales of the SDQ was re-identified, confirming the original structure of the instrument [9].

The internal consistency was satisfactory to good for each of the five subscales of the SDQ in the total sample and in each individual country participating in the study. For the total sample, the internal consistency was highest for the SDQ subscales measuring prosocial behaviour, peer problems and conduct problems. These results are comparable to those obtained with a child psychiatric sample [1] and demonstrate that the basic psychometric properties of the SDQ are stable across countries and clinical samples.

As expected, it was apparent from the present SDQ data set that children with ADHD had more problems than children of a random epidemiological sample in the UK. This holds true for all SDQ scales and underlines the need to perform general screening rather than just evaluating ADHD core symptoms [8].

Age effects were observed for some subscales of the SDQ parent-reported version. More hyperactivity-inattention symptoms and more peer problems were reported for younger as compared to older children. Younger children with ADHD were also

generally more impaired than older children (SDQ total difficulties score). Other studies have indicated that SDQ scores are to some extent age-dependent [9].

In this study, a significant *gender effect* demonstrated that girls had more emotional problems and showed more prosocial behaviour than boys. There were no gender differences for the conduct problems, hyperactivity-inattention, peer problems and total difficulties scores. These findings are not surprising, since the same tendency was indicated by previous studies [9].

A sound and useful psychometric instrument should be applicable in cross-cultural settings by showing either similarities (e.g. to collapse data sets) or differences (e.g. to evaluate relevant cultural factors). We found that country had a significant effect on each of the SDQ subscale scores. For the SDQ hyperactivity-inattention subscale, children from France and the UK were rated as being most impaired, whereas children from Switzerland and Iceland had the lowest scores. However, as the samples from France and the UK were much larger than those from Switzerland and Iceland, it is difficult to draw firm conclusions about potential cultural differences.

Prüß et al. [5] suggested that the setting of the inquiry may have effects on the ratings obtained in the psychopathological assessment of children. However, in our study, no significant differences were found for the factor *"investigator type"*, indicating that the health professional seen did not influence the results of the parent-rated SDQ. Thus, this potential bias may not be relevant when testing for the central hypotheses of ADORE.

Finally, the SDQ subscales and total difficulties score had varying associations with the other scales used in the ADORE study. The highest correlation was seen between the SDQ conduct problems subscale and the Risk Avoidance domain of the CHIP-CE. These scales probably measure similar aspects of behaviour, with less risk

avoidance leading to more conduct problems. Moreover, a moderate correlation was found between the ADHD-RS-IV scores and the total difficulties score of the SDQ. This may reflect psychosocial impairment in ADHD. Furthermore, it is in line with a consistent relationship between the SDQ hyperactivity-inattention subscale and the hyperactivity-impulsivity subscale of the ADHD-RS-IV, and the high association between the emotional symptoms subscale of SDQ and a lack of satisfaction and comfort as measured by the CHIP-CE.

In conclusion, the present results demonstrate that the SDQ parent ratings of children with ADHD provide relevant clinical information. Furthermore, the SDQ has shown adequate psychometric properties indicating that the obtained data sets may be used for further statistical analyses within the planned longitudinal design of the ADORE study.

References

1. Becker A, Woerner W, Hasselhorn M, Banaschewski T, Rothenberger A (2004) Validation of the parent and teacher SDQ in a clinical sample. European Child and Adolescent Psychiatry 13 (suppl. 2): 11–16

2. Döpfner M, Steinhausen H-C, Coghill D, Dalsgaard S, Poole L, Preuss U, Ralston SJ, Rothenberger A on behalf of the ADORE study group (2006) Cross-cultural reliability and validity of ADHD assessed by the ADHD Rating Scale in a pan-European study. European Child and Adolescent Psychiatry 15 (supp. 1): 46–55

3. Gillberg C, Gillberg IC; Rasmussen P, Kadesjö B, Söderström H, Rastam M, Johnson M, Rothenberger A, Niklasson L (2004) Co-existing disorders in ADHD

– implications for diagnosis and intervention. European Child and Adolescent Psychiatry 13 (supp. 1): 80–92

4.   Preuss U, Ralston SJ, Baldursson G, Falissard B, Lorenzo MJ, Pereira RR, Vlasveld L, Coghill D on behalf of the ADORE study group (2006) Study design, baseline patient characteristics and intervention in a cross-cultural framework: results from the ADORE study. European Child and Adolescent Psychiatry 15 (supp. 1): 4–14

5.   Prüß U, von Widdern S, von Ferber Ch, von Ferber L, Lehmkuhl G (2005) Do different survey settings influence the prevalence of symptoms? A methodological comparison using the youth self-report. Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie 33:295–305

6.   Ralston SJ, Lorenzo MJM and the ADORE study group (2004) ADORE – Attention-Deficit Hyperactivity Disorder Observational Research in Europe. European Child and Adolescent Psychiatry 13 (supp. 1): 36–42

7.   Riley AW, Coghill D, Forrest CB, Lorenzo MJ, Preuss U, Ralston SJ, Spiel G and the ADORE study group (2006) Validity of the health-related quality of life assessment in the ADORE study: parent report form of the CHIP-Child Edition. European Child and Adolescent Psychiatry 15 (supp. 1): 63–71

8.   Rothenberger A, Danckaerts M, Döpfner M, Sergeant J, Steinhausen HC (2004) EINAQ – A European educational initiative on Attention-Deficit Hyperactivity Disorder and associated problems. European Child and Adolescent Psychiatry 13 (supp. 1): 31–35

9.   Rothenberger A, Woerner W (2004) Strength and Difficulties Questionnaire (SDQ) – Evaluations and Applications. European Child and Adolescent Psychiatry 13 (suppl. 2): 1–2 [editorial]

10. Taylor E, Döpfner M, Sergeant J, Asherson P, Banaschewski T, Buitelaar J, Coghill C, Danckaerts M, Rothenberger A, Sonuga-Barke E, Steinhausen HC, Zuddas A (2004) European clinical guidelines for hyperkinetic disorder – first upgrade. European Child and Adolescent Psychiatry 13 (supp. 1): 7–30

Table 1: Factor structure and internal consistency of the parent-rated SDQ (N =1,459)**

Extracted factor :	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Internal consistency (α)
Prosocial behaviour						
1 Considerate	**0.57**					
4 Shares	**0.62**					**0.72**
9 Caring	**0.71**					
17 Kind to kids	**0.57**					
20 Helps out	**0.73**					
Peer problems						
6 Solitary		**0.67**				
11 Good friend *		**0.58**				**0.72**
14 Popular *		**0.69**				
19 Picked on, bullied		**0.62**				
23 Best with adults		**0.68**				
Emotional symptoms						
3 Somatic complaints			**0.49**			
8 Worries			**0.71**			**0.66**
13 Unhappy			**0.63**			
16 Clingy			**0.62**			
24 Fears			**0.65**			
Conduct problems						
5 Tempers				**0.46**		
7 Obedient *				**0.37**		**0.69**
12 Fights	0.36			**0.54**		
18 Lies, cheats				**0.75**		
22 Steals				**0.69**		
Hyperactivity-Inattention						
2 Restless					**0.69**	
10 Fidgety					**0.70**	**0.58**
15 Distractible					**0.56**	
21 Reflective *					**0.45**	
25 Persistent *					**0.54**	

* Before analyses, the value of the items were reversed. ** Only rotated loadings ≥ 0.35 are shown

Table 2: Scale means by gender and age for parent-rated SDQ scores

SDQ Subscale	Total sample	UK Norms[1]	Younger age(6-10 years)	Older age (11-18 years)	Age effects	Age effects (two-tailed)	Boys	Classified as abnormal[2], %	Girls	Classified as abnormal[2], %	Gender effects	Gender effects (two-tailed)
	(N=1,453)	(N=10,298)	(N=1,110)[a]	(N=334)[a]	F-value*	p	(N=1,222)		(N=231)		F-value*	p
Total difficulties score	20.4 (6.0)	8.4 (5.8)	20.5 (6.0)	19.7 (5.9)	5.89	0.0154	20.4 (6.0)	73.1	20.4 (6.1)	73.5	0.27	0.6029
Emotional symptoms	4.0 (2.4)	1.9 (2.0)	3.9 (2.4)	4.0 (2.5)	1.78	0.1825	3.9 (2.4)	39.7	4.5 (2.4)	49.1	14.14	0.0002
Conduct problems	4.5 (2.3)	1.6 (1.7)	4.5 (2.3)	4.3 (2.3)	2.84	0.0920	4.5 (2.3)	62.8	4.2 (2.3)	60.0	1.48	0.2242
Hyperactivity-Inattention	8.3 (1.7)	3.5 (2.6)	8.3 (1.7)	8.1 (1.8)	10.01	0.0016	8.3 (1.7)	85.6	8.0 (2.0)	77.0	3.77	0.0525
Peer problems	3.7 (2.5)	1.5 (1.7)	3.8 (2.5)	3.3 (2.4)	12.13	0.0005	3.7 (2.5)	48.9	3.7 (2.4)	53.5	0.00	0.9852
Prosocial behaviour	6.8 (2.3)	8.6 (1.6)	6.8 (2.3)	6.9 (2.2)	0.74	0.3888	6.7 (2.3)	16.2	7.3 (2.1)	9.6	17.40	<0.0001

Data are presented as mean (SD) unless indicated otherwise; [a]N values for the age subgroups do not total N=1,453 due to missing data; UK = United Kingdom; *degrees of freedom=1

[1]For parent SDQ for sample aged 5-15 years; available at www.sdqinfo.com

[2]Abnormal defined as a score ≥ 17 for total difficulties, ≥ 5 for emotional symptoms, ≥ 4 for conduct problems, ≥ 7 for hyperactivity-inattention, ≥ 4 for peer problems and ≤ 4 for prosocial behaviour

Table 3: Parent-rated SDQ scores by country (N=1,459)

	Parent-rated SDQ scores					
	Total difficulties	**Emotional symptoms**	**Conduct problems**	**Hyperactivity - Inattention**	**Peer problems**	**Prosocial behaviour**
All Countries (N=1,459)	20.4 (6.0)*	4.0 (2.4)*	4.5 (2.3)	8.3 (1.7)	3.7 (2.5)	6.8 (2.3)
Austria (N=71)	20.7 (5.8)**	3.9 (2.3)**	4.6 (2.2)	8.3 (1.6)	3.9 (2.5)	6.8 (2.1)
Denmark (N=31)	21.3 (6.0)	4.8 (2.3)	3.7 (2.1)	8.7 (1.4)	4.0 (2.7)	7.2 (2.1)
France (N=241)	22.6 (5.8)	4.5 (2.5)	5.2 (2.2)	8.7 (1.3)	4.3 (2.5)	7.1 (2.6)
Germany (N=425)	19.3 (5.9)	4.0 (2.3)	4.1 (2.2)	7.7 (1.8)	3.5 (2.5)	6.8 (2.1)
Iceland (N=47)	17.1 (6.4)	3.5 (2.7)	3.4 (2.0)	7.5 (2.1)	2.8 (2.1)	6.9 (2.0)
Italy (N=109)	20.0 (5.3)	3.2 (2.0)	4.9 (2.0)	8.6 (1.5)	3.3 (2.2)	7.0 (2.1)
Netherlands (N=207)	19.5 (5.7)	3.9 (2.4)	3.9 (2.4)	8.5 (1.6)	3.3 (2.4)	7.2 (2.1)
Norway (N=50)	18.9 (5.1)	3.3 (2.2)	3.4 (1.6)	8.6 (1.7)	3.6 (2.6)	6.9 (2.4)
Switzerland (N=56)	17.6 (6.6)	3.8 (2.5)	3.6 (2.1)	7.2 (2.0)	3.0 (2.4)	7.5 (1.6)
UK (N=222)	22.5 (5.7)	3.9 (2.5)	5.6 (2.5)	8.9 (1.5)	4.0 (2.2)	5.9 (2.3)

Data are presented as mean (SD); *N=1458; **N=70
Higher scores reflect more difficulties except for prosocial behaviour where a higher score reflect more strengths

Table 4: Correlation coefficients between parent-rated SDQ scores and scores from other scales used in the ADORE study (N=1,315)[a]

	Parent-rated SDQ scores					
	Total difficulties	**Emotional symptoms**	**Conduct problems**	**Hyperactivity-Inattention**	**Peer problems**	**Prosocial behaviour**
Clinician Rated						
ADHD-RS-IV						
Total score	0.50	0.15	0.42	0.51	0.31	-0.22
Inattention subscale score	0.36	0.17	0.28	0.30	0.23	-0.18
Hyperactivity-impulsivity subscale score	0.47	0.09	0.42	0.54	0.29	-0.20
CGI-S score	0.38	0.20	0.29	0.24	0.29	-0.17
CGAS score	-0.34	-0.18	-0.29	-0.16	-0.27	0.20
Parent Rated						
CHIP-CE domain scores						
Satisfaction	-0.46	-0.50	-0.25	-0.06	-0.36	0.30
Comfort	-0.54	-0.57	-0.35	-0.20	-0.30	0.14
Resilience	-0.37	-0.18	-0.36	-0.15	-0.29	0.41
Risk Avoidance	-0.63	-0.22	-0.73	-0.38	-0.38	0.40
Achievement	-0.60	-0.32	-0.41	-0.26	-0.58	0.35

ADHD-RS-IV, clinician scored parent version of ADHD-Rating Scale-IV; CGI-S, Clinical Global Impression-Severity; CGAS, Children's Global Assessment Scale; CHIP-CE, Child Health and Illness Profile-Child Edition; $p \leq .001$ for all correlations
[a]Includes only those patients with scores for the SDQ and the other scales measured

5. DISCUSSION

5.1 Cumulative research from various cultures

Although the studies reviewed here represent only a selection among many standardization, validation and application studies which have already been published, they provide a basis for empirical evidence of the psychometric characteristics of the SDQ versions in the various cultures and study contexts. In a great variety of cultures and languages, the experience gathered using the SDQ supports European evidence of good psychometric properties and clinical utility of this questionnaire.

However, there are considerable differences between the scale values obtained and the cut-off values for each norming sample. Whether the resulting differences reflect the heterogeneous sample composition in relation to age, gender or socioeconomic status cannot be determined in the context of this review. In the adaptations process, certain items may be more in need of cultural-specific adaption than others (van Widenfelt, 2005). For example, Wang and Ollendick (Wang, 2001) discovered that in Chinese culture there is no equivalent term for "self-esteem" as defined in Western cultures.

Furthermore, it is not possible to determine the true reasons underlying minor differences in mean SDQ scores since even after excluding sampling bias, age and gender differences as well as improper translation as potential sources, these deviations between normative data from two countries could reflect either real differences in children's behaviour or different standards (and tolerance levels) applied by raters when they report on the child's behaviour. The deviations may even result from very subtle differences in the wording of the items which, although being equivalent and correct translations, could affect the informant's readiness to endorse

a particular statement. In large samples, as usually required and examined for calibration and standardization purposes, even very small effects can easily attain statistical significance. Such confounded factors are extremely difficult to disentangle, thus underlining the necessity of establishing national norms, but also posing an additional challenge to adequate interpretations of cross-cultural comparisons (Woerner, 2004b).

In this context it should be elucidated why children's behaviour in different cultures is evaluated so differently by their parents and teachers (given the same prevalence for child psychiatric disorders). Are these children really more disturbed behaviourally than their peers in another country or does their rating by parents or teacher rather reflect a more strict set of criteria for child behaviour in a culture in which conspicuous behaviour is less tolerated?

A problem or symptom may be more pathological or more familial in one culture versus another. Cultural beliefs can influence the raters view of the acceptability of individual behaviour or characteristics as well as what type of interaction and relationship are acceptable (Rubin, 1998).

Mojtabai concluded in his work (Mojtabai, 2005) that: "The cross-country differences in the prevalence of different emotional and behavioural problems need to be corroborated in future studies using other instruments and, ideally, multiple informants. Future studies also need to examine the implications of cross-country differences in mental health care for quality of care, nature of treatments, satisfaction with services, and the course and outcome of childhood psychopathology in the examined countries".

Since worldwide usage of the SDQ can be expected to increase in the future, more international coordination is encouraged in order to fully exploit the promising

potentials of this versatile assessment tool and systematically investigate cross-cultural differences and similarities in child and adolescent behaviour. Whether re-standardization can conceal real cross-cultural differences in the prevalence of mental health problems still remains to be clarified. Further clinical studies are necessary to investigate whether the SDQ is suitable for discriminating samples of children and adolescents from two or more different countries who have undergone psychiatric treatment.

One of the advantages of the SDQ is its impact supplement providing an estimate of burden, which is a part of the diagnostic criteria in child and adolescent psychiatry. It is probable that a combination of symptom and impact scores will be the best indicator of caseness also in other countries, as was the case for the British version of the SDQ (Goodman, 2000b) and the use of the extended version of the SDQ is recommended in future studies.

5.2 Comprehensive discussion of the four SDQ studies

After a comprehensive discussion of the SDQ studies, the most important results of the four articles described above should be mentioned briefly. In the first article on the validation of the parent and teacher vesion of the SDQ it could be shown that the parent and teacher SDQs proved to be valid and helpful questionnaires for use in the framework of a multi-dimensional behavioural assessment, and appear to be well-suited for screening purposes, longitudinal monitoring of therapeutic effects, and scientific research purposes. Furthermore, it was shown that the self-rated version of the SDQ was a reliable and valid method for the assessment of behavioural problems in children and adolescents. In the absence of adult informant reports from parents and teachers, the diagnostic value of self-ratings was also demonstrated. In a further study it was demonstrated that the SDQ subscale hyperactivity/attention

deficit shows very good discriminatory validity for differentiating between patients with and without ADHS. Thus, the SDQ subscale hyperactivity/attention deficit provides a helpful building block for diagnosing ADHS in a multidimensional diagnostic framework. However, a significant value for this subscale should not be seen as evidence for a diagnosis of ADHS and additional diagnostic measures should be undertaken.

Finally, another prospective pan-European study with ADHD children confirmed the good validity and reliability of the parent-reported SDQ scale structure and showed that the scale scores are dependent on age and gender. In contrast to investigator type, different cultures had a significant effect on SDQ scores. Correlations with other scales used in the ADORE study underline both separate domains and meaningful associations.

6. REFERENCES

Achenbach, T. M. (1991). Manual for the child behavior checklist /4-18 and 1991 profile. University of Vermont, Department of Psychiatry: Burlington.

Almaqrami, M. H., & Shuwail, A. Y. (2004). Validity of the self-report version of the strengths and difficulties questionnaire in Yemen. Saudi Med J, 25 (5), 592-601.

Ambrosini, P. J. (2000). Historical development and present status of the schedule for affective disorders and schizophrenia for school-age children (K-SADS). J Am Acad Child Adolesc Psychiatry, 39 (1), 49-58.

Ashenafi, Y., Kebede, D., Desta, M., & Alem, A. (2001). Prevalence of mental and behavioural disorders in ethiopian children. East Afr Med J, 78 (6), 308-311.

Association, A. P. (1994). Diagnostic and statistical manual of mental disorders, 4th edition (DSM-IV). Washington DC: American Psychiatric Association.

Banaschewski, T., Woerner, W., Becker, A., & Rothenberger, A. (2004). Diagnostik der Aufmerksamkeitsdefizit-Hyperaktivitäts-Störung: Unterstützung durch den Elternfragebogen zu Stärken und Schwächen des Kindes. Monatsschrift Kinderheilkunde, 151 (12), 778-781.

Becker, A., Hagenberg, N., Roessner, V., Woerner, W., & Rothenberger, A. (2004a). Evaluation of the self-reported SDQ in a clinical setting: Do self-reports tell us more than ratings by adult informants? Eur Child Adolesc Psychiatry, 13 Suppl 2, 17-24.

Becker, A., Roessner, V., Rothenberger, A., & Banaschewski, T. (2006a). Anwendung des SDQ-Elternfragebogens für Screening und Diagnosestellung einer ADHS. Kinderärztliche Praxis, 77 (4), 230-235.

Becker, A., Steinhausen, H. C., Baldursson, G., Dalsgaard, S., Lorenzo, M. J., Ralston, S. J., et al. (2006b). Psychopathological screening of children with ADHD: Strengths and Difficulties Questionnaire in a pan-european Study. Eur Child Adolesc Psychiatry, 15 Suppl 1, 56-62.

Becker, A., Woerner, W., Hasselhorn, M., Banaschewski, T., & Rothenberger, A. (2004b). Validation of the parent and teacher SDQ in a clinical sample. Eur Child Adolesc Psychiatry, 13 Suppl 2, 11-16.

Bourdon, K., Goodman, R., Rae, D., Simpson, G., & Koretz, D. (2005). The Strengths and Difficulties Questionnaire: U.S. Normative data and psychometric properties. Journal of the American Academy of Child & Adolescent Psychiatry, 44 (6), 557-564.

Conners, C. K. (1998). Rating scales in attention-deficit/hyperactivity disorder: Use in assessment and treatment monitoring. J Clin Psychiatry, 59 Suppl 7, 24-30.

Dickey, W. C., & Blumberg, S. J. (2004). Revisiting the factor structure of the strengths and difficulties questionnaire: United states, 2001. J Am Acad Child Adolesc Psychiatry, 43, 1159-1167.

Fisher, P. W., Shaffer, D., Piacentini, J. C., Lapkin, J., Kafantaris, V., Leonard, H., et al. (1993). Sensitivity of the diagnostic interview schedule for children, 2nd edition (disc-2.1) for specific diagnoses of children and adolescents. J Am Acad Child Adolesc Psychiatry, 32 (3), 666-673.

Fleitlich, B., & Goodman, R. (2001). Social factors associated with child mental health problems in brazil: Cross sectional survey. British Medical Journal, 323, 599-600.

Gillberg, C. (1998). Hyperactivity, inattention and motor control problems: Prevalence, comorbidity and background factors. Folia Phoniatr Logop, 50 (3), 107-117.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. J Child Psychol Psychiatry, 38 (5), 581-586.

Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. J Am Acad Child Adolesc Psychiatry, 40 (11), 1337-1345.

Goodman, R., Ford, T., Corbin, T., & Meltzer, H. (2004). Using the Strengths and Difficulties Questionnaire (SDQ). Multi-informant algorithm to screen looked-after children for psychiatric disorders. Eur Child Adolesc Psychiatry, 13 Suppl 2, 25-31.

Goodman, R., Ford, T., Richards, H., Gatward, R., & Meltzer, H. (2000a). The development and well-being assessment: Description and initial validation of an integrated assessment of child and adolescent psychopathology. J Child Psychol Psychiatry, 41 (5), 645-655.

Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000b). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. Br J Psychiatry, 177, 534-539.

Goodman, R., Meltzer, H., & Bailey, V. (1998). The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. Eur Child Adolesc Psychiatry, 7 (3), 125-130.

Goodman, R., & Scott, S. (1999). Comparing the Strengths and Difficulties Questionnaire and the child behavior checklist: Is small beautiful? J Abnorm Child Psychol, 27 (1), 17-24.

Green, H., McGinnity, A., Meltzer, H., Ford, T., & Goodman, R. (2005). Mental health of children and young people in great britain 2004. London: DH Publications.

Hawes, D. J., & Dadds, M. R. (2004). Australian data and psychometric properties of the Strengths and Difficulties Questionnaire. Aust N Z J Psychiatry, 38 (8), 644-651.

Heubeck, B. G. (2000). Cross-cultural generalizability of cbcl syndromes across three continents: From the USA and holland to australia. J Abnorm Child Psychol, 28 (5), 439-450.

Ickowicz, A., Schachar, R. J., Sugarman, R., Chen, S. X., Millette, C., & Cook, L. (2006). The parent interview for child symptoms: A situation-specific clinical research interview for Attention-Deficit Hyperactivity and related disorders. Can J Psychiatry, 51 (5), 325-328.

Kashala, E., Elgen, I., Sommerfelt, K., & Tylleskar, T. (2005). Teacher ratings of mental health among school children in kinshasa (Democratic Republic of Congo). Eur Child Adolesc Psychiatry, 14 (4), 208-215.

Klasen, H., Woerner, W., Rothenberger, A., & Goodman, R. (2003). German version of the Strengths and Difficulties Questionnaire (SDQ-german)-Overview and evaluation of initial validation and normative results. Prax Kinderpsychol Kinderpsychiatr, 52 (7), 491-502.

Koskelainen, M., Sourander, A., & Kaljonen, A. (2000). The Strengths and Difficulties Questionnaire among finnish school-aged children and adolescents. Eur Child Adolesc Psychiatry, 9 (4), 277-284.

Liang, H., Flisher, A. J., & Chalton, D. O. (2002). Mental and physical health of out of school children in a south african township. Eur Child Adolesc Psychiatry, 11 (6), 257-260.

Malhotra, S., & McElroy, S. L. (2002). Medical management of obesity associated with mental disorders. J Clin Psychiatry, 63 Suppl 4, 24-32.

McCarty, C. A., Weisz, J. R., Wanitromanee, K., Eastman, K. L., Suwanlert, S., Chaiyasit, W., et al. (1999). Culture, coping, and context: Primary and secondary control among thai and american youth. J Child Psychol Psychiatry, 40 (5), 809-818.

Meltzer, H., Gatwood, R., Goodman, R., & Ford, T. (2000). Mental health of children and adolescents in great britain. London: Mortality Statistics, Office for National Statistics.

Mojtabai, R. (2005). Trends in contacts with mental health professionals and cost barriers to mental health care among adults with significant psychological distress in the united states: 1997-2002. Am J Public Health, 95 (11), 2009-2014.

Mojtabai, R. (2006). Serious emotional and behavioral problems and mental health contacts in american and british children and adolescents. J Am Acad Child Adolesc Psychiatry, 45 (10), 1215-1223.

Mullick, M. S., & Goodman, R. (2001). Questionnaire screening for mental health problems in Bangladeshi children: A preliminary study. Soc Psychiatry Psychiatr Epidemiol, 36 (2), 94-99.

Muris, P., Meesters, C., Eijkelenboom, A., & Vincken, M. (2004). The self-report version of the strengths and difficulties questionnaire: Its psychometric properties in 8- to 13-year-old non-clinical children. Br J Clin Psychol, 43 (4), 437-448.

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. Eur Child Adolesc Psychiatry, 12 (1), 1-8.

Obel, C., Heiervang, E., Rodriguez, A., Heyerdahl, S., Smedje, H., Sourander, A., (2004). The Strengths and Difficulties Questionnaire in the nordic countries. Eur Child Adolesc Psychiatry, 13 Suppl 2, 32-39.

Organization, W.-W. H. (1992). The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines. Geneva: WHO.

Quay, H. C. (1986). Psychopathological disorders of childhood: John Wiley & Sons.

Rothenberger, A., & Woerner, W. (2004). Strengths and Difficulties Questionnaire (SDQ)-Evaluations and applications. Eur Child Adolesc Psychiatry, 13 Suppl 2, 1-2.

Rubin, K. H. (1998). Social and emotional development from a cultural perspective. Dev Psychol, 34 (4), 611-615.

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). Int J Methods Psychiatr Res, 14 (3), 158-166.

Simpson, G. A., Bloom, B., Cohen, R. A., Blumberg, S., & Bourdon, K. H. (2005). U.S. Children with emotional and behavioral difficulties: Data from the 2001, 2002, and 2003 national health interview surveys. Adv Data (360), 1-13.

Sims, A. (2003). Symptoms in the mind: An introduction to descriptive psychopathology (3 rd ed.). Edinburgh: Elsevier Science Ltd.

Smedje, H., Broman, J. E., Hetta, J., & von Knorring, A. L. (1999). Psychometric properties of a swedish version of the " Strengths and Difficulties Questionnaire ". Eur Child Adolesc Psychiatry, 8 (2), 63-70.

Taylor, E., Schachar, R., Thorley, G., & Wieselberg, M. (1986). Conduct disorder and hyperactivity: Separation of hyperactivity and antisocial conduct in british child psychiatric patients. Br J Psychiatry, 149, 760-770.

Thabet, A. A., Stretch, D., & Vostanis, P. (2000). Child mental health problems in arab children: Application of the Strengths and Difficulties Questionnaire. Int J Soc Psychiatry, 46 (4), 266-280.

van Widenfelt, B. M., Treffers, P. D., de Beurs, E., Siebelink, B. M., & Koudijs, E. (2005). Translation and cross-cultural adaptation of assessment instruments used in psychological research with children and families. Clin Child Fam Psychol Rev, 8 (2), 135-147.

Vostanis, P. (2006). Strengths and Difficulties Questionnaire: Research and clinical applications. Curr Opin Psychiatry, 19 (4), 367-372.

Wang, Y., & Ollendick, T. H. (2001). A cross-cultural and developmental analysis of self-esteem in chinese and western children. Clin Child Fam Psychol Rev, 4 (3), 253-271.

Woerner, W., Becker, A., Friedrich, C., Klasen, H., Goodman, R., & Rothenberger, A. (2002). Normal values and evaluation of the german parents' version of Strengths and Difficulties Questionnaire (SDQ): Results of a representative field study. Z Kinder Jugendpsychiatr Psychother, 30 (2), 105-112.

Woerner, W., Becker, A., & Rothenberger, A. (2004a). Normative data and scale properties of the german parent SDQ. Eur Child Adolesc Psychiatry, 13 Suppl 2, 3-10.

Woerner, W., Fleitlich-Bilyk, B., Martinussen, R., Fletcher, J., Cucchiaro, G., Dalgalarrondo, P. (2004b). The Strengths and Difficulties Questionnaire overseas: Evaluations and applications of the SDQ beyond europe. Eur Child Adolesc Psychiatry, 13 Suppl 2, 47-54.

ACKNOWLEDGMENTS

Zuerst möchte ich mich ganz herzlich bei Dr. Wolfgang Woerner bedanken, der mich geduldig in das wissenschaftliche Arbeiten eingeführt hat und dafür sorgte, dass diese Arbeit zustande gekommen ist. Er war lange Zeit mein Wegbegleiter und Kollege. Danke für die vielen hilfreichen Stunden zur Planung, Gestaltung und Realisierung im Rahmen dieses Projektes. Wenn auch sein Ziel nun ein anderes ist, so teilten und teilen wir doch die gleichen Sorgen und Freuden auf diesem Weg.

Herzlich bedanken möchte ich mich auch bei Prof. Dr. Marcus Hasselhorn, der es immer wieder verstand, zu motivieren und die Arbeit schließlich zu einem guten Ende führte. Ebenso gilt mein herzlicher Dank Prof. Dr. Rothenberger für die Möglichkeit, in der Kinder- und Jugendpsychiatrie Göttingen dieses Projekt durchzuführen und für die konstruktiven Vorschläge zur Gestaltung dieser Arbeit.

Mein Dank gilt allen Kolleginnen und Kollegen, die mich auf verschiedene Weise unterstützt und ermutigt haben. Für die emotionale Unterstützung in Zeiten mit einer "erhöhten Erlebnisdichte" möchte ich mich ebenso bei Edgar Bunker bedanken, der es verstanden hat, die Dinge immer wieder in einem hoffnungsvollen Licht erscheinen zu lassen.

Mein herzlichster Dank geht zum Schluss natürlich auch an meine Familie. Patricia, Danke,…dass du nie müde geworden bist, meine Prioritätenliste immer wieder auf den neusten Stand zu bringen…und an meine Mutter, die ebenso stoisch an ein erfolgreiches Ende meiner Arbeit geglaubt hat und mich in Zeiten drohender Mangelernährung zu versorgen wußte.

Andreas Becker   Göttingen, den 19.03.2007

LIST OF PUBLICATIONS

Woerner, W., Becker, A., Friedrich, C., Klasen, H., Goodman, R., Rothenberger, A. (2002). Normierung und Evaluation der deutschen Elternversion des Strengths and Difficulties Questionnaire (SDQ): Ergebnisse einer repräsentativen Felderhebung. Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie 30 (2), 105-112.

T. Banaschewski, W. Woerner, A. Becker und A. Rothenberger (2004). Diagnostik der Aufmerksamkeitsdefizit-Hyperaktivitäts-Störung: Unterstützung durch den Elternfragebogen zu Stärken und Schwächen des Kindes (SDQ) Monatsschrift Kinderheilkunde 151 (12), 370-377

V. Roessner, T Banaschewski, H Uebel, A Becker, A Rothenberger (2004). Neuronal network models of ADHD-lateralization with respect to interhemispheric connectivity reconsidered. ECAP Supplement 1.

Wolfgang Woerner, Andreas Becker, Aribert Rothenberger (2004). Normative data and scale properties of the German parent SDQ. Eur Child and Adolesc Psychiatry 13 (2)

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

A. Becker, N. Hagenberg, V. Roessner, W. Woerner, A. Rothenberger (2004).
Evaluation of the self-reported SDQ in a clinical setting: Do self-reports tell us more than ratings by adult informants? European Journal of Child and Adolescent Psychiatry 13 (2) 17-23.

A. Becker & Aribert Rothenberger, Deutsche ADORE/FACE Studiengruppe (2006).
Psychopathological screening of children with ADHD: Strengths and Difficulties Questionnaire in a paneuropean study. European Journal of Child and Adolescent Psychiatry 15 (1) 56-62.
A. Becker, Veit Rössner, Aribert Rothenberger, Tobias Banaschewski (2006). Anwendung des SDQ-Elternfragebogens für Screening und Diagnosestellung einer ADHS. Kinderärztliche Praxis 77 (4) 230-235.

Roessner V, Becker A, Banaschewski T, Rothenberger A. (2005). Tic disorders and obsessive compulsive disorder: where is the link? J Neural Transm Suppl. 2005;(69):69-99

Roessner V, Becker A, Banaschewski T, Rothenberger A. (2006). Psychopathological Profile in Children with Chronic Tic Disorder and Co-existing ADHD: Additive Effects. J Abnorm Child Psychol. (pub ahead.)

Tobias Banaschewski , Sinje Ruppert , Rosemary Tannock , Björn Albrecht , Andreas Becker , Henrik Uebel , Joseph A Sergeant , Aribert Rothenberger (2006). Colour perception in ADHD. J Child Psychol Psychiatry. 47 (6):568-572.

Peter Müller, Andreas Becker und Georg Stolpmann (2006). Vergleich externer und interner Prognose-Gutachten im Maßregelvollzug Sachsen-Anhalts. Recht & Psychiatrie (4): 174-180

Roessner V, Uebel H, Becker A, Beck G, Bleich S, Rothenberger A: Serum level of semicarbazide-sensitive amine oxidase in children with ADHD. Behavioral and Brain Function 2:5.

Roessner V, Weber A, Becker A, Beck G, Frieling H, Bleich S: Decreased SSAO serum levels in patients with major depression. Progress in Neuro-pychoharmacology and Biological Psychiatry 30: 906-909.

Veit Roessner, Annette Weber, Andreas Becker, Georg Beck, Helge Frieling, Stefan Bleich (2007). Decreased serum activity of semicarbazide-sensitive amine oxidase (SSAO) in patients treated with second generation antipsychotics: a link to impaired glucose metabolism? European Journal of Clinical Pharmacology (pub ahead.)

Borwin Bandelow, Ulrich Seidler-Brandler, Andreas Becker, Dirk Wedekind, Eckart Rüther (2007). Meta-Analysis of Randomized Controlled Comparisons of Psychopharmacological and Psychological Treatments for Anxiety Disorders. The World Journal of Biological Psychiatry (pub ahead.).

Veit Roessner, Andreas Becker, Tobias Banaschewski, Aribert Rothenberger, Roger Freeman (2007). Developmental psychopathology of children and adolescents with Tourette Syndrome – impact of ADHD. European Journal of Child and Adolescent Psychiatry (pub ahead.)

Veit Roessner, Andreas Becker, Aribert Rothenberger (2007). Psychopathologie bei Tic- und Zwangsstörungen. Kindheit und Entwicklung (pub ahead.)

Veit Roessner, Andreas Becker, Tobias Banaschewski, Aribert Rothenberger (2007). Executive functions in children with chronic tic disorders with/without ADHD - new insights. European Journal of Child and Adolescent Psychiatry (pub ahead.)

Thomas Achenbach, Andreas Becker, Manfred Döpfner, Einar Heiervang, Veit Roessner, Hans-Christoph Steinhausen, Aribert Rothenberger (2007). Measurement of Psychopathology - CBCL and SDQ in a cross-cultural and globalizing framework. European Journal of Child and Adolescent Psychiatry (pub ahead.)

CURRICULUM VITAE

Persönliche Daten

Name: Andreas Becker

geboren am: 07.12.1966 in Göttingen

Familienstand: verheiratet

Staatsangehörigkeit: deutsch

Schul- und Berufsausbildung

1973 – 1977	Grundschule in Ebergötzen
1977 – 1983	Realschule, Abschluß mit erweiterter mittlerer Reife, Göttingen
1983 – 1984	Berufsgrundbildungsjahr Physik-Biologie-Chemie in Göttingen
1984 – 1986	Ausbildung zum Physiklaboranten am Max-Planck-Institut für biophysikalische Chemie in Göttingen
1986 – 1987	Fachoberschule Technik, Abschluß mit Fachhochschulreife, Göttingen
1987 – 1989	Fachgymasium Technik, Abschluß mit Abitur, Göttingen

Hochschulausbildung

1989 – 1991	Physikstudium an der Universität Göttingen
1991 – 1993	Psychologiestudium an der Universität Göttingen, Diplomvorprüfung
1993 – 2000	Psychologiestudium an der Universität Göttingen, Abschluß mit Diplom

Fachspezifische Fortbildungen und Tätigkeiten

1994 – 1999	Weiterbildung in Gesprächstherapie und ehrenamtliche Mitarbeit in der Telefonseelsorge, Göttingen
1996 – 1999	Weiterbildung in Verhaltenstherapie, Zwischenbescheinigung der DGVT
1997 – 2000	Studentische Hilfskraft an der Klinik und Poliklinik für Kinder- und Jugendpsychiatrie der Universität Göttingen
1999 – 2008	Durchführung forensischer Begutachtungen an der Psychiatrischen Klinik der Universität Göttingen
2007 – 2008	Durchführung familienrechtlicher Begutachtungen an der Kinder- und Jugendpsychiatrie der Universität Göttingen

2000 - 2001   wissenschaftliche Mitarbeit (40 Std./ Monat) an der Klinik und
                     Poliklinik für Kinder- und Jugendpsychiatrie Göttingen.
                     Aufgabenbereich: Entwicklung und Organisation eines Verfahrens zur
                     Dokumentation von Patientendaten

2002 - 2005   wissenschaftliche Mitarbeit (80 Std./ Monat) an der Klinik und
                     Poliklinik für Kinder- und Jugendpsychiatrie Göttingen.
                     Aufgabenbereich: Forschung & Lehre

2005 - 2008   Ausbildungsleitung an der Pflegeschule für Krankenschwestern in
                     Kinderpsychiatrie und allgemeiner Psychiatrie an der Universität
                     Göttingen

2006 - 2008   wissenschaftliche Mitarbeit (Vollzeit) als Forschungsleiter an der
                     Klinik und Poliklinik für Kinder- und Jugendpsychiatrie Göttingen.
                     Aufgabenbereich: Forschung, Lehre & Patientenversorgung

Göttingen, den 03.03.2008

Andreas Becker