Abstract: Background: Discrepancies between multiple informants often create considerable uncertainties in delivering services to youth. The present study assessed the ability of the parent and youth scales of the Strength and Difficulties Questionnaire (SDQ) to predict mental health problems/disorders across several mental health domains as validated against two contrasting indices of validity for psychopathology derived from the Development and Well Being Assessment (DAWBA): (1) an empirically-derived computer algorithm and (2) expert based ICD-10 diagnoses. Method: Ordinal and logistic regressions were used to predict any problems/disorders, emotional problems/disorders and behavioural problems/disorders in a community sample (n=252) and in a clinic sample (n=95). Results: The findings were strikingly similar in both samples. Parent and youth SDQ scales were related to any problem/disorder. Youth SDQ symptom and impact had the strongest association with emotional problems/disorder and parent SDQ symptom score were most strongly related to behavioural problems/disorders. Both the SDQ total and the impact scores significantly predicted emotional problems/disorders in males whereas this was the case only for the total SDQ score in females. Conclusion: The present study confirms and expands previous findings on parent and youth informant validity. Clinicians should include both parent and youth for identifying any mental health problems/disorders, youth information for detecting emotional problems/disorders, and parent information to detect behavioural problems/disorders. Not only symptom scores but also impact measures may be useful to detect emotional problems/disorders, particularly in male youth.
Title: “The contribution of parent and youth information to identify mental health disorders or problems in adolescents”.
Journal: Child and Adolescent Psychiatry and Mental Health

Dear Editors,

Thank you for considering the above mentioned manuscript for publication in Child and Adolescent Psychiatry and Mental Health. We thank the reviewers for their detailed comments and suggestions. We revised the paper accordingly. Below you find a point-to-point response to the reviewers’ comments.

Reviewer #1: The authors have been able to address the concerns of the reviewers; the rationale for the study is more clear now; the presentation of results has improved; the discussion is richer.

Reviewer #2: These alterations have improved reporting of methods and results, but there are still a few issues:
Are hypotheses 3 and 4 a priori? Because they aren’t in the original manuscript, and if they are a posteriori, it should be stated.

Response: Thank you for these suggestions. We agree and revised the text in the following way: Hypotheses 3 and 4 have been established a posteriori in accordance with findings from previous studies.

Is the CI for parent SDQ total score predicting expert diagnosis of any problem/disorder correct?
Response: The CI for parent SDQ total score predicting expert diagnosis of any problem/disorder has been corrected (0.11-1.27 instead of 1.11-1.27).

Address the question of how low correlation between both diagnostic methods could affect the validity of your results, especially regarding emotional disorders on the community sample.
Response: The correlations between DAWBA bands and expert diagnosis are in a moderate range (see table 2) for most disorders/problems but somewhat lower for emotional problems/disorders (r=.31). Whereas DAWBA bands are based on computer algorithms, expert diagnoses also consider further information based on open-end text fields. Previous studies also found low correlations between clinically established and interview based diagnoses, in particular for emotional disorders (e.g. Rettew et al. 2009). We added a short comment on the lower correlations in the discussion section. “Based on rather low prevalence rates of affective and anxiety disorders the corresponding correlations of DAWBA bands and expert ratings were only modest in the community sample. This finding may also reflect the rather moderate agreements of different diagnostic approaches when assessing affective and anxiety disorders in youth [44].”

Reviewer #3: Improvements in outlining core ideas in many respects. The authors are getting close to a structure that lays out their ideas and findings comparing self-report questionnaires compared to a validation standard. Some of the edits, however, struggled with grammar and language that distracted from overall flow/clarity. If I’m understanding the flow of the manuscript correctly, self-report screening measures are commonly used but struggle with cross-informant agreement. Literature offers various reasons for different informant reports, but existing models for reconciling multiple informant report is difficult. SDQ scales were compared to DAWBA bands and diagnoses to see how parental and youth self-report compared to these diagnostic indices. Based on these data, diagnoses better identified by parental report, youth report, or both are discussed. If I’m understanding that much correctly, then I’m hoping the following considerations might add some clarity:

Hypotheses offered on page 6-7. Consider referral back to these hypotheses in Discussion to complete the thought and help organize the paper.
Response: We revised the discussion and now refer to the former hypotheses.

Page 10, lines 259-265. I'm assuming these data are from the ICD-10 expert ratings. Would help to just simply state that so the reader doesn't have to think to differentiate which information came from DAWBA bands vs. DAWBA expert diagnoses vs. SDQ etc.

Response: We added this information.

Page 11, lines 268-269. If SDQ comparisons to DAWBA results are being used to identify meaningful ways to look at cross-informant inconsistency, what does it mean if SDQ total score and impact in the clinical sample are not significantly correlated to DAWBA results. It seems this should be acknowledged and discussed in the Discussion (such as end of page 13).

Response: The findings on page 11, lines 268-269 refer to correlations between parent and youth SDQ scores (table 3) but not to correlations between SDQ and DAWBA measures. SDQ total and impact scores from parents and youths were not significantly correlated. We added a sentence in the discussion regarding the missing correlation of parent and youth SDQ and impact scores.

Page 11, lines 277-288. Consider separating discussion of problem bands vs. ICD-10 diagnoses. It just requires as fair amount of concentration to keep track of which variables are being compared. I was confused at first why sometimes authors were referencing "problems" and other times "diagnoses." Or consider some clearer indicator of DAWBA bands vs. diagnoses.

Response: We added more information at the beginning of the sections concerning DAWBA bands and expert diagnoses as outcomes.

In introduction, authors note problems with past studies' validation measures such as "relying on clinician's diagnoses of unclear reliability." It might help just to clarify why (a) DAWBA bands and diagnoses service as good indices for validation and (b) just noting in a sentence how this approach improves on past methodology problems. This is just a suggestion, but hopefully helps to tie up loose ends from the introduction to discussion. I appreciate that you moved away from using the term "gold standard" but the question remains regarding why DAWBA is a reasonable measure for comparison.

Response: We added a statement on the advantages of the DAWBA as a multi-informant based outcome measure: "Because the DAWBA is a well validated multi-informant based instrument, the current study may overcome some methodological limitations with diagnoses derived from single informants or unstructured clinical evaluations".

Page 16, Strengths and limitations. What potential impact on clinical utility should be considered given the study inclusion criteria that aimed for "minimization of the ethnic heterogeneity" of the sample?

Response: We added the following limitation in the revised manuscript. "Because the community sample was based on European ethnicities, the findings may not generalize to other ethnic groups".

Various details and grammar:
Still inconsistent use of "e.g.," vs. "e.g." and same for "i.e.," vs. "i.e."
Please review for hyphenation use and consistency. For example, "self-report" while other times using "self report." Also consider terms like "empirically-developed" (page 6, line 152).
Please review for consistency in use of serial comma or not.
Page 5, line 123. "and conduct disorders (ODD)," - first of all, should be "(CD)". Also, for consideration, should these references to ADHD, ODD, and CD be singular vs. plural?
Page 6, line 154. Not sure the semicolon makes sense here.
Page 7, line 179. IMAGEN study - does IMAGEN stand for something? Should that full
Response: Thank you for your careful reading. We corrected these mistakes according to the suggestions. There is no full name for the IMAGEN study.

Reviewer #4: Altogether, the revised paper now has a much more straightforward and focused structure. The added information contributed well to understanding the research questions and the conclusions made from the analyses. The analyses now are better suited for the complex data structure regarding distribution of frequencies of dependent variables in the community sample. There only some minor remarks.

INTRODUCTION

The introduction section now has a concise and straightforward structure focusing on the information that is required to understand the research questions. The sentence "In community samples, adolescents described themselves as having considerably more behavioral problems than their parents [18, 24] and ..." (p.5, l.127) sounds misleading, although the intended meaning is clear.

Response: We reformulated this sentence in the following way: "In community samples, adolescent self-reports show higher levels of behavioural problems than parent reports [18, 24]..."

Maybe it would further clarify the measurement procedure, if the authors would call the youth SDQ score a self report measure.

Response: We decided to further use the term "youth SDQ" (vs. parent SDQ) in agreement with previous publications. We clarify in the method section that the youth SDQ is a self-report instrument.

One question emerges from the last paragraph of the introduction: Is it in accordance with scientific practice to report information about the distribution of the sample of the study (low prevalence of certain disorders) already in the introduction section? Perhaps the authors should formulate a more general analysis strategy here and describe the limitation regarding the community sample in the results section.

Response: The assumption that prevalence rates of specific disorders are rather low in community samples is not specific to the present study. The information that some analyses were only feasible in the clinic sample may help the reader to better understand the current findings.

METHOD

p.7, l. 182: "... based ON two criteria: ...

Response: Revised accordingly.

It would have been helpful to describe if the two criteria of sample recruitment were achieved with the sample by reporting socio-economic background variables.

Response: We do not have specific information on socio-economic background
variables in the Mannheim sample which is a subsample of the IMAGEN sample. Further information on the sample composition of the IMAGEN study has been provided in previous publications (e.g. Schumann et al 2010).

The SDQ is well established. However, did the authors consider computing and reporting any psychometric properties?

Response: We added the following sentence in the method part: “The psychometric properties of the SDQ are well established [1, 39] so that we did not compute them again in the present study.”

RESULTS

In the results section the application of Firth’s logit is an appropriate solution for the problem of handling low sample sizes in logistic regression.

On page 10 in the first paragraph about results, the structure of the sentences needs a correction. In addition, the single behavioral disorders do not sum up to the complete number of behavioral disorders.

Response: We revised the paragraph accordingly. The number of behavioural disorders in the clinic sample is 21 instead of 22. We corrected this finding and recalculated the regression analyses (and found similar results).

Looking at the correlations between parent reports and youth self-reports, two aspects seem a little bit confusing: First, the correlations of the peer problem subscale is missing, and second, the correlations between all reported subscales is highly significant in the clinical sample, but the total score is not. Please explain.

Response: We only included parent-youth correlations of the SDQ total and impact score as well as the relevant subscales (behaviour problems, emotional problems, hyperactivity) for this study. The peer-problems and the pro-social subscale of the SDQ were not considered because these scales are not specific to psychiatric disorders. Thus, they were not included in the regression analyses. We added the following information to the discussion: However, the correlations between all reported subscales were highly significant in the clinical sample, but the total score was not. There is no clear and easy explanation to this sample-dependent finding that is in need of more detailed studies.

DISCUSSION

Perhaps the discussion would profit from the report and discussion of the main findings at the beginning, and the report of main findings start with page 14. The discussion about the validation approaches, which now starts the discussion section is important. However, it may structurally be better to put this part after reporting the main findings. The paragraph on p.16 l.404 repeats on page 17. Please delete one of them.

Response: Thank you for this suggestion. However, we decided not to change the structure of the discussion part. We report the main findings after repeating the study aims and discussing the descriptive findings. However, we now refer to the hypotheses and added some headings in the discussion for a better guidance of the readers. We deleted the repeats on page 17.

We think that the revised manuscript is now well suited for publication in CAPMH.

Yours sincerely,
Marcel Aebi, corresponding author

References


The contribution of parent and youth information to identify mental health disorders or problems in adolescents

Marcel Aebi1,2,3, marcel.aebi@uzh.ch

Christine Kuhn1, christine.kuhn@puk.zh.ch

Tobias Banaschewski4, tobias.banaschewski@zi-mannheim.de

Yvonne Grimmer4, Yvonne.Grimmer@zi-mannheim.de

Luise Poustka5, Luise.Poustka@med.uni-goettingen.de

Hans Christoph Steinhausen1,6,7, hc.steinhausen@kjpd.uzh.ch

Robert Goodman8, robert.goodman@kcl.ac.uk

1 Department of Child and Adolescent Psychiatry, University Hospital of Psychiatry Zurich, Switzerland
2 Department of Forensic Psychiatry, University Hospital of Psychiatry Zurich, Switzerland
3 Clinical Psychology for Children/Adolescents and Couples/Families, Department of Psychology, University of Zurich, Switzerland
4 Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Germany
5 Department of Child and Adolescent Psychiatry/Psychotherapy, University of Göttingen, Germany
6 Child and Adolescent Mental Health Centre, Capital Region Psychiatry, Copenhagen, Denmark
7 Clinical Psychology and Epidemiology, Department of Psychology, University of Basel, Switzerland
8 Department of Child and Adolescent Psychiatry, King’s College London Institute of Psychology, Psychiatry & Neuroscience, United Kingdom

Correspondence to
Marcel Aebi
University Hospital of Psychiatry
Neptunstrasse 60, 8032 Zurich
Switzerland
Phone: +41 43 556 40 13
Fax: +41 43 556 40 05
E-Mail: marcel.aebi@uzh.ch
Abstract

Background: Discrepancies between multiple informants often create considerable uncertainties in delivering services to youth. The present study assessed the ability of the parent and youth scales of the Strength and Difficulties Questionnaire (SDQ) to predict mental health problems/disorders across several mental health domains as validated against two contrasting indices of validity for psychopathology derived from the Development and Well Being Assessment (DAWBA): (1) an empirically-derived computer algorithm and (2) expert based ICD-10 diagnoses. Method: Ordinal and logistic regressions were used to predict any problems/disorders, emotional problems/disorders and behavioural problems/disorders in a community sample (n=252) and in a clinic sample (n=95). Results: The findings were strikingly similar in both samples. Parent and youth SDQ scales were related to any problem/disorder. Youth SDQ symptom and impact had the strongest association with emotional problems/disorder and parent SDQ symptom score were most strongly related to behavioural problems/disorders. Both the SDQ total and the impact scores significantly predicted emotional problems/disorders in males whereas this was the case only for the total SDQ score in females. Conclusion: The present study confirms and expands previous findings on parent and youth informant validity. Clinicians should include both parent and youth for identifying any mental health problems/disorders, youth information for detecting emotional problems/disorders, and parent information to detect behavioural problems/disorders. Not only symptom scores but also impact measures may be useful to detect emotional problems/disorders, particularly in male youth.

Keywords: adolescent psychopathology; emotional problems, behavioural problems, multi-informants; SDQ; DAWBA
Introduction

Youth and parent screening measures such as the Strength and Difficulties Questionnaire [SDQ; 1, 2] or the Achenbach Systems of Empirically based assessments [ASEBA; 3] are easy to use and cost-effective methods to identify adolescents with psychological difficulties. Both of these instruments are highly popular among mental health practitioners and researchers and also among other child care professionals. They have been translated into many different languages and implemented in clinical processes worldwide. Mental health professionals use these screening measures to decide whether further and more detailed assessments of emotional or behavioural disorders are indicated. Researchers use these screening measures in epidemiological and clinical studies to measure the type, the extent, and the course of mental health problems. Nurses and practitioners in general hospitals and social workers in schools and juvenile justice institutions use these screening measures to decide which adolescents need more specific assessment and treatment and should be referred to mental health practitioners. However, discrepancies between multiple informants often create considerable uncertainties in delivering services to youth and drawing conclusions from research [4].

Informant discrepancies on mental health problems are one of the major challenges in child and adolescent psychiatry. A recent meta-analysis of 341 studies [5] found that modest cross-informant agreement is one of the most robust phenomena in clinical child and adolescent research (with mean correlation: $r=.28$). However, the degree of cross-informant agreement for mental disorders varies between mental health domains, different societies and cultures and also depends on the youth’s age and gender [5-8].

A number of different factors contribute to informant discrepancies on mental health problems [9, 10]. First, some mental health problems emerge only in specific situations such as school and family contexts or within peer interactions. Contextual variations occur within a variety of
psychiatric domains including social anxiety, attention-deficit-hyperactivity, and conduct

Secondly, informants (e.g., parent and youth) may differ on their perceptions and awareness of mental health problems and what kinds of behaviours are within the norm. For example, parents may be worried about the adolescent’s withdrawal, whereas the adolescent perceives his behaviour as within the normal range and views the intrusiveness of the parents as the area of concern. Thirdly, informant discrepancies may result from measurement errors in regard to the frequency and severity of behavioural, emotional or hyperactivity problems.

Different strategies have been suggested for how to choose informants and how to aggregate data from multiple informant data for diagnostic decision making [12, 14]. In order to disentangle three meaningful components of psychopathology such as (1) the trait (measure of interest for youth’s psychopathology), (2) the context (factors related to the emergence and the reporting of symptoms), and (3) the informants perspective, principal component analysis and regression analyses have been proposed [15, 16]. However, these approaches are quite complex and cannot easily be implemented into clinical practice.

Two factors seem crucial for researchers and clinicians to decide whether parent or youth information is more accurate: 1) the area of mental health problems addressed (e.g., emotional vs. behavioural problems) and 2) the context in which the assessment took place (e.g., clinical vs. community assessments) [17, 18]. For detecting any mental health problems, information from both informants can be useful [19]. In a community sample, parent and youth information uniquely and indispensably contributed to later signs of maladjustment (referral to mental health services, need for professional help, and presence of a disorder) [20]. Similarly, both self-reports and parent reports were found necessary to detect the presence of a psychiatric diagnosis in a clinical outpatient sample [17].
To explore emotional problems/disorders such as depression and anxiety, clinicians and researchers usually rely on adolescents’ self-reports from questionnaires or interviews because adolescents themselves are assumed to be the most valid source of information for this kind of problems [21]. In fact, adolescents do report significantly more internalizing symptoms than their parents in clinical samples [22, 23] and community samples [24]. Furthermore, self-information has been found accurate to predict the presence of internalizing problems/emotional disorders in community as well as in clinical samples [8, 17, 20, 21, 25-27]. However, some studies also found that the inclusion of parent information further increased the ability to detect emotional problems in community and clinical samples [17, 28].

In the exploration of externalizing problems such as attention-deficit-hyperactivity disorder (ADHD), oppositional defiant disorder (ODD), and conduct disorder (CD), parent information has been considered to be more valid than youth self-reports by mental health professionals [21]. Though on theoretical grounds, self-reports also seem important to assess conduct problems, because many of these behaviours (e.g., thefts, fire setting, physical attacks) occur in setting to which parents are not privy [22]. In community samples, adolescent self-reports show higher levels of behavioural problems than parents reports [18, 24] and adolescent self-reports were found to be valid predictors of externalizing problems, behavioural disorders and later criminal behaviours [20, 28-31]. In clinical samples, adolescents may underreport behavioural problems [18, 32] and adolescent self-reports are sometimes less accurate than parent reports in detecting behavioural disorders [17]. Some adolescents may minimize their conduct problems to avoid possible adverse consequences of full disclosure [33].

Previous studies testing the informant validity of parent and adolescent self-ratings reported conflicting findings and were limited by the use of either just community or just clinical
samples and by a paucity of validation measures, (e.g., relying on clinicians’ diagnoses of unclear reliability). Furthermore, previous studies did not consider impact measures as additional information to detect psychiatric disorders. Some adolescents find it hard to report psychological symptoms and may find it easier to describe specific impairments in school, family and peer group. Given the previous findings on the validity of the SDQ impact scales, we predicted that impact measures in addition to symptoms scores would make a useful contribution to the assessment of mental health disorders.

The present study intended to confirm and expand previous findings by analysing data collected in a community and an outpatient sample. The ability of parent and youth SDQ scales measuring problems and impact were analysed in order to predict mental health problems/disorders across several mental health domains (any disorder, emotional disorders, behavioural disorders), as validated against two contrasting indices of validity derived from the Development and Well-Being Assessment, DAWBA (see method section below): One approach used the empirically-developed multi-informant DAWBA bands (ordinal measures) based on a computer algorithm to aggregate parent and/or youth information from structured interview questions, while the other approach used ICD-10 diagnosis generated by expert DAWBA raters, i.e., experienced clinicians who rated the presence of an ICD-10 disorders after reviewing the answers to closed and open-ended questions. Because the DAWBA is a well validated multi-informant based instrument [35, 36], the current study may overcome some methodological limitations of diagnoses derived from single informants or unstructured clinical evaluations.

Based on the existing literature, we hypothesized that in multivariate analyses 1) the youth and parent SDQ total scores would both be highly associated with any problems/disorders in both samples, 2) the youth SDQ total score would be more highly associated with emotional problems/disorders than the parent SDQ total score in both samples, 3) parent and youth SDQ
total scores would be associated with behavioural problems/disorders in the community sample, but only parent SDQ total score would be associated with behavioural problems/disorders in the clinical sample. Hypotheses 3 and 4 were established a posteriori in accordance with findings from previous studies. We further assumed that youth and parent SDQ impact scores would supplement the predictive power of symptoms scores in the prediction of any problems/disorders, emotional problems/disorders, and behavioural problems/disorders in both samples.

In addition, we tested the ability of the SDQ conduct and emotional problem scales in the prediction of emotional and behavioural problems/disorders in both samples. Further supplemental analyses of parent and youth SDQ hyperactivity and conduct problem scales in the prediction of ODD, CD and ADHD were performed in the clinic sample only (because of the low prevalence rates of these disorders in the community sample).

Method

Samples

The present study is based on a community and clinic sample from two different sites [19]. The community sample is one arm of the IMAGEN study described in more detail in [37]. A sample of healthy adolescents was recruited from secondary schools in the city of Mannheim, Germany, and surrounding areas via flyers, school visits and residents’ registration offices. The recruitment was based on two criteria: 1. Greatest possible diversity in terms of socio-economic status, cognitive and emotional development. To achieve this goal, private- and state-funded schools and special educational schools (classes) were equally targeted; 2. Minimization of the ethnic heterogeneity by selecting a sample of young people with European ethnicity. Exclusion criteria were severe complications during pregnancy and birth, serious pre-existing conditions,
(particularly neurological and psychiatric disorders), IQ < 70 and contraindications for a parallel magnetic resonance imaging study, e.g., severe claustrophobia or metal/electronical implants [37]. The study was approved by the local ethics committee of the University of Mannheim. The final community sample consisted of 252 adolescents (46.8% male) with a mean age of 13.98 years (SD=0.60 years, range 13-17 years).

The outpatient sample was recruited from all available patients who attended the outpatient centres of the child and adolescent psychiatry service of the canton of Zurich, Switzerland, between September 2007 and June 2009 (n=875). Out of this sample, 345 youth and parents with sufficient German language skills participated (participation rate = 40.5%). However, only patients aged 11 to 17 years with available parent and youth information were considered for the present study. There were no further exclusion criteria [35]. The final outpatient sample consisted of 95 patients (66.3% male) with a mean age of 13.95 years (SD=2.04 years, range 11-17 years). Subjects in both the community and clinical samples were first assessed with the internet-based parent and youth versions of the SDQ [2, 38] and then filled in the online version of the Development and Well-Being Assessment [DAWBA; 36].

Measures

**Strength and Difficulties Questionnaire (SDQ):** The SDQ is a questionnaire covering common mental health problems in children aged 2 to 17. The 20 items relating to emotional symptoms, conduct problems, hyperactivity and peer problems can be summed to generate a total difficulty score ranging from 0 to 40. The SDQ has been shown to have dimensional as well as categorical qualities [1]. The SDQ is commonly administered with an impact supplement that asks whether the respondent thinks the youth has significant difficulties, and if so inquires about overall distress and social impairment – forming the basis for an impact score.

In this study, **the parent and self-report versions** of the SDQ with impact supplement was
administered to parents and to youths aged 11 or older and used as a screening measure to predict DAWBA bands/expert ratings across multiple mental health domains. The psychometric properties of the SDQ are well established [1, 39] so that we did not compute them again in the present study.

Development and Well-Being Assessment (DAWBA): The DAWBA [36] includes structured interview sections covering the major mental disorders, followed by a semi-structured part eliciting open-ended descriptions from respondents about areas of concern. Diagnostic predictions in line with ICD-10 and DSM-IV criteria can be generated by computerized algorithms drawing on data from the structured questions, generating what are called “DAWBA bands” [40]. The DAWBA bands are based on an algorithm that combines the information from symptom and impact measures from all available respondents, e.g., parent report and adolescent report. It is not an average or an addition, but aims to follow the logic of the DSM and ICD classifications, e.g., giving more weight to symptoms of hyperactivity if reported across different situations and accompanied by impairment. The DAWBA bands algorithm does not prioritise any one category of informant a priori. DAWBA bands have been previously validated in two large samples of British (n=7912) and Norwegian youth (n=1364) [40]. In the present study we use the “any disorder” DAWBA band, the emotional disorder DAWBA band (affective and anxiety problems) and the behavioural disorder DAWBA band. Supplemental analysis also included specific DAWBA bands for ADHD, ODD, and CD) Since the DAWBA bands are quick, cheap and standardized [40], they have been used as the only source of diagnostic ratings in some research studies [e.g., 41]. The DAWBA bands are used as ordinal outcome measures in the present study (frequencies of the probability to meet criteria of a disorder: <0.5%, ~3%, ~15%, ~50%, 70%+). In addition, dichotomous (present versus absent) ratings of ICD-10 disorders (emotional, behavioural, ADHD, CD and ODD) were generated by expert clinicians based on a review of all available information, including open-ended
The inter-rater reliability for expert based diagnosis was found to be good (kappa 0.79-0.89) [35].

Statistical analyses

We used multivariate ordinal and logistic regression to predict total, emotional, behavioural and ADHD DAWBA bands/disorders. Besides z-transformed SDQ youth and parent symptom and impact scores we included youth’s age and male gender (males = 1, females = 0) as covariate in the analyses. Because of the small number of psychiatric disorders in the community sample, Firth's bias-reduced logistic regressions by the use of the package "logistf" [42] in R statistical software were performed [43]. This method is accurate for logistic regression analyses with rare outcome data. None of the linear predictors/covariates showed multicollinearity and the assumption of proportional odds was met for all ordinal regression analyses (chi2 > 0.05). In addition, sex-specific receiver operating characteristic (ROC) analyses of SDQ total and impact scores were performed to predict DAWBA expert rated emotional disorders. All analyses were conducted using R statistical software [43] and SPSS 23 for Mac OS X, were two-tailed, and utilized a threshold for statistical significance of p = 0.05.

Results

Frequencies of the DAWBA bands of the 252 adolescents of the community and the 95 adolescents of the clinic sample are shown in table 1. As expected and in contrast to the clinical sample, most adolescents from the community sample showed low probabilities for having a mental health disorder according to DAWBA expert ratings (e.g., 3% and less, table 1). In the community sample 21 (8.3%) adolescents had any ICD-10 disorder, 6 (2.4%) any emotional disorder, 9 (3.6%) any behavioural disorder (ODD: 1, 0.4%; CD: 8, 3.2%), and 6 (2.4%) any hyperkinetic disorder. In the clinic sample 67 (70.5%) adolescents had any ICD-10 disorder, 41
(43.2%) any emotional disorder, 21 (22.1%) any behavioural disorder (ODD: 13, 13.7%; CD: 8, 8.4%), and 13 (13.7%) any hyperkinetic disorder. Bivariate correlations of DAWBA bands and disorders (expert diagnosis) in the community and clinical samples are shown in table 2. All correlations were in the medium range and highly significant in both samples. Bivariate correlations between parent and youth SDQ scores and subscales in the community and the clinical sample are presented in table 3. With the exception of the SDQ total score and SDQ impact in the clinic sample, all correlations were in the medium range and highly significant in both samples.

Findings in the community sample

Multivariate ordinal and Firth’s bias reduced logistic regressions with DAWBA bands (problems) and expert diagnoses (disorders) as outcome variables are presented in table 4 and show that the parent SDQ total score (but not the impact score) was related to any problems and disorders, any behavioural problems and disorders, but not to any emotional problems or disorders. The youth SDQ total score was associated with any problems as well as to emotional problems and disorders. The youth SDQ impact score was related to any problems and disorders as well as to emotional problems. Among the SDQ subscales, the parent SDQ emotional problems scale was associated with emotional problems but not with emotional disorders, whereas the youth SDQ emotional problems scale was associated with emotional problems and disorders. The parent but not the youth SDQ behaviour problems subscale was related to any behaviour problems and disorders. Among the covariates, age was negatively related to the presence of an emotional disorder (coefficient=-2.54, 95%CI=-4.97-
Findings in the clinic sample

Findings from multivariate ordinal and logistic regressions with DAWBA bands (problems) and expert diagnoses (disorders) as outcome variables are presented in table 5. The parent SDQ total score (but not the impact score) was related to any problems as well as to behavioural problems and disorders. The youth SDQ total score was associated with any problems and disorders as well as with emotional disorders. The youth SDQ impact score was related to emotional problems. The SDQ emotional problems subscales were related to emotional problems and disorders, particularly in the youth report, and to a lesser degree in the parent report. The parent SDQ behaviour problems subscale was associated with behavioural problems and disorders. The youth SDQ behaviour problem subscale was related to a lesser degree than the parent SDQ behaviour problems scale to behavioural problems only. Among the covariates female gender was significantly associated with the presence of an emotional disorder (OR=2.90, 95%CI= 1.05-8.05) and male gender with the presence of a behavioural disorders (OR=0.12, 95%CI= 0.02-0.66).

Findings based on supplemental analyses in the clinic samples for specific problems/disorders are presented in table S1. The parent SDQ total score was related to hyperactivity problems, conduct problems and disorders, and oppositional problems and disorders, whereas the youth SDQ total score was not related to any of these scales. Neither the parent nor the youth SDQ impact scale was associated with any of these problems/disorders. The parent SDQ
hyperactivity scale was related to hyperactivity problems and disorders and the parent SDQ behaviour problems was related to conduct problems and disorders as well as to oppositional defiant problems and disorders. The youth SDQ behaviour problems scale was associated with conduct problems only.

Finally, additional ROC analyses (with the area under the curve (AUC) as a measure of diagnostic accuracy) in the clinic sample found that both the SDQ total (AUC=.71, 95%CI=.59-.84, p=0.004) and the impact score (AUC=.67, 95%CI=.52-.83, p=0.025) were significantly associated with emotional disorder in male youth. Interestingly, the SDQ impact score had higher sensitivity values whereas the total score had higher specificity values (see figure 1). In female youth, only the SDQ total score (AUC=.75, 95%CI=.56-.93, p=0.024) but not the impact score (AUC=.58, 95%CI=.37-.78, p=0.487) was significantly related to emotional disorders.

Discussion

The current study adds to previous findings on the validity of multi-informant assessments of mental disorders in youth [5, 19]. Unlike earlier studies, the present investigation is based on internet-based instruments only. The DAWBA has previously been used to identify mental health disorders with similar properties to traditional diagnostic interviews such as the Diagnostic Interview Schedule for Children (DISC) and the Child and Adolescent Psychiatric Assessment (CAPA) [44]. However, the DAWBA was a more conservative measure, generating fewer diagnoses than the other two measures [44]. In the present study, two different approaches to validation were used in parallel across multiple mental health domains: First, validation against an empirically-derived computerized algorithm (the DAWBA bands) and,
secondly, validation against ICD-10 diagnoses by clinical experts. Overall, the two validation approaches generated similar results supporting the likely robustness of the findings. Based on the rather low prevalence rates of affective and anxiety disorders, the corresponding correlations of DAWBA bands and expert ratings were only modest in the community sample. This finding may also reflect the rather moderate agreement of different diagnostic approaches when assessing affective and anxiety disorders in youth [45]. Correlation coefficients between parent and youth SDQ scales were similar to findings from previous studies [6, 7]. However, the correlations between all reported subscales were highly significant in the clinical sample, but the total score was not. There is no clear and easy explanation to this sample-dependent finding that is in need of more detailed studies. Furthermore and in contrast to our and previous findings in community samples [34], youth and parents in the clinic sample did not agree on the level of distress and impairment caused by mental health problems. Also this finding needs further studies aiming at some clarification of the origins of these discrepant views.

**Parent and youth information to identify any mental health problems/disorders**

Our findings confirmed and expanded previous findings on informant validity in both community and clinical samples of youth, [e.g., 22, 46]. In line with previous research and in agreement with hypothesis 1, we found that both parent and youth information is valuable for identifying mental health problems in adolescents. Each category of informant made its own unique and valuable contribution to the prediction of mental health problems in both community and clinical settings. Therefore, researchers and clinicians are strongly recommended to collect information from both youth and parents whenever possible for assessing mental health problems [19], though parent reports alone are sometimes a reasonable substitute for screening purposes when it would be impractical or unaffordable to collect information from multiple informants.
Also in agreement with previous research and in confirming hypothesis 2, we found that youth self-reports are the best source for identifying emotional problems such as depression and anxiety in adolescents. The superiority of self-reports was independent of sample characteristic and therefore may apply for researchers assessing prevalence rates in the community as well as for practitioners in psychiatric institutions. One of the reasons is that parents may have limited access to youth’s intrapsychic processes. The superiority of self-report may not apply to younger children under the age of 11, who may not have the ability to describe their emotional problems. Furthermore, our results as well as findings of previous research show that parent information can still significantly add value for diagnostic decision making and problem description. Future screening instruments may use different sets of items for parent and youth to address internalizing disorders. Parent scales should specifically focus on observable behaviours that are associated with depression and anxiety (e.g., social isolation, avoidance behaviours).

Independent of the setting (clinical vs. community sample), we found parent reports better suited than youth self-reports for identifying behavioural problems/disorders and specifically for CD and ODD in adolescents. According to hypothesis 4, our findings confirm results of previous studies based on clinical settings that adolescent self-report show limited value for assessing ADHD [46, 47], CD [48], and ODD [32, 49]. Although some studies have previously found higher correlations between parent and youth reports for externalizing disorders [5-7, 19] and that self-reports can discriminate youth referred for conduct disorder from normal controls [50], our findings show limited additional value resulting from including self-reports to detect externalizing mental health problems in both the community and clinical samples. In clinical settings, youth may minimize problems to gain favorable reports from their clinicians. Some
youth may be repressing and denying their behavioral problems or providing socially desirable responses in questionnaires [33]. In community samples, self-reports have previously been found useful in screening for externalizing disorders [20, 28-31]. Our results do not confirm these findings and hypothesis 3 and are in keeping with a clinical body of opinion that adolescent information only is not sufficient to decide on behaviour problems/disorders.

Furthermore, and supporting the need for multi-informant data, parent-reported behavior problems in community youth outperformed adolescent self-reports in the prediction of later criminal outcomes in adolescence and adulthood [31]. However, given the limited sample size and the low prevalence of behaviour disorders/problems in our community study, the present findings should be treated with caution.

The value of impact measures for identifying mental health problems/disorders

Most previous studies have focused on the presence of mental health symptoms only, rather than on how these symptoms influence individual, family and school functioning [34]. The present findings support the relevance of the youth SDQ impact score for detecting emotional problems in male adolescents in clinical settings and for detecting mental health problems/disorders in community youth. Some youth may report subclinical levels of symptoms but still report distress and impairments caused by these problems. Previous research found subclinical symptoms of adolescent depression to have long term negative effects in adulthood [51]. Our findings may indicate that the SDQ impact scale is useful for screening of early mental health problems. Our additional ROC analyses provided some indication of gender-specific differences in the identification of emotional disorders in the clinic sample. Anxious or depressed males who do not report much by way of emotional symptoms may nevertheless be aware that their life is impaired. If clinicians ask about such impairment and follow up with sensitive probing about emotional symptoms, this might improve the recognition of anxiety and depression, particularly in males.
**Strengths and limitations**

This is the first study that has tested parent and youth screening measures comprehensively across multiple mental health domains simultaneously in clinical and community settings with two complementary approaches to validation (empirically-validated computer algorithms and diagnoses by expert clinicians). It is reassuring that the results of the two approaches converge, supporting informant-specific assessment of psychopathology in youth. Nevertheless, the present findings have to be interpreted under the view of some limitations: First, because of the moderate sample size of the clinic sample and the low prevalence of some disorders, the statistical power for the regression analyses was limited. We therefore only provided analyses for the most frequent disorders. Secondly, the present findings were limited to the SDQ as predictor and the DAWBA as outcome. No further screening measures of psychopathology were used in the present study. Thirdly, no teacher ratings were available and could therefore not be included as further informants in this study. Forthly, because the community sample was based on European ethnicities, the findings may not generalize to other ethnic groups. Finally, family background variables (e.g., socio-economic status or parental separation) were not available and could not have been controlled for in the present study. Further studies are needed to elucidate the underlying mechanisms of discrepancies of informant validity.

**Conclusions**

The current findings illustrate the importance of considering motivation and the nature of behavioural and emotional problems in self-report ratings. Clinical practitioners should keep in mind that adolescents may display problem behaviours only in specific settings but also have limited ability to report behavioural and hyperactivity problems. The “Operations Triad Model” [OTM; 5, 10] is a conceptual framework on how to use and interpret multi-informant assessments which is guided by evidence based information on the divergence and convergence of informants’ reports. OTM guides clinicians a) to hypothesize about patterns of convergence
and divergence among informants reports and b) to develop personalized assessments that
directly test these hypotheses. To do this, practitioners may rely on information on the context
in which the problems emerge as well as the informant’s ability to report mental health
problems across different domains. The current findings may guide clinicians to choose which
kind of information should be collected from which informants and how to aggregate that
information in order to decide on further assessment and treatment.

List of abbreviations

SDQ = Strength and Difficulties Questionnaire; DAWBA = Development and Well
Being Assessment; ICD-10 = International Classification of Diseases, Tenth Edition; ADHD =
Attention Deficit Hyperactivity Disorders; CD = Conduct Disorders; ODD = Oppositional
Defiant Disorders (ODD); ASEBA = Achenbach Systems of Empirically based assessments;
DSM-5 = Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; SD = Standard
Deviation; ROC = Receiver Operating Characteristic; DSM-IV = Diagnostic and Statistical
Manual of Mental Disorders, Forth Edition; SPSS = Statistic Package for Social Scientists;
AUC = Area under the curve.

Declarations

Ethics approval and consent to participate

The Zurich clinical study was approved by the local ethics committee of the Canton of
Zürich and is registered as a randomized clinical trial (ISRCTN19935149). The Mannheim
study was approved by the local ethics Committee of the University of Mannheim. All
participants agreed either to participate in the Zurich or Mannheim study.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analysed during this study are included in this published article
Competing interests

Robert Goodman is owner of Youthinmind Ltd, which produces no-cost and low-cost websites related to the SDQ and DAWBA. Tobias Banaschewski served in an advisory or consultancy role for Hexal Pharma, Lilly, Medice, Novartis, Otsuka, Oxford outcomes, PCM scientific, Shire and Viforpharma. He received conference attendance support and conference support or received speaker’s fee by Lilly, Medice, Novartis, and Shire. He is/has been involved in clinical trials conducted by Lilly, Shire, and Viforpharma. The present work is unrelated to the above grants and relationships. During the last three years, Hans-Christoph Steinhausen has been a speaker for Shire Pharmaceuticals and received book royalties from Cambridge University Press, Elsevier, Hogrefe, Huber, Klett, and Kohlhammer publishers. The present work is unrelated to the above mentioned grants and relationships. All other authors report no conflict of interests with the present study.

Funding

There was no external funding of the Zurich study. The Mannheim sample is one arm of the IMAGEN study that received funding from the EU Commission in FP6.

Authors’ contributions

MA and RG were responsible for the basic conceptualization of the article, conducted the statistical analyses and wrote the manuscript. HCS, RG and CK were responsible for the design and the data collection and management of the original Zurich study and TB, YG and LP were responsible for the data collections and management of the Mannheim arm of the IMAGEN study. HCS and CK made substantial contributions to the final manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Christa Winkler Metzke from the Department of Child and Adolescent Psychiatry,
University Hospital of Psychiatry Zurich, Switzerland who helped with the data collocation and the DAWBA diagnostic ratings of the Zurich sample.
References


32. Loeber R, Green SM, Lahey BB, Stouthamer-Loeber M. Differences and similarities


40. Goodman A, Heiervang E, Collishaw S, Goodman R. The 'DAWBA bands' as an


Table 1: Frequencies of probands in the community (n=252) and the clinic sample (n=95) according to probability of having any disorder, any emotional, any behavioural disorder, ADHD, CD and ODD (DAWBA bands)

<table>
<thead>
<tr>
<th>DAWBA bands</th>
<th>“Any DAWBA”</th>
<th>Emotional DAWBA</th>
<th>Behavioural DAWBA</th>
<th>ADHD DAWBA</th>
<th>CD DAWBA</th>
<th>ODD DAWBA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of having a disorder</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community</td>
<td>Clinic</td>
<td>Community</td>
<td>Clinic</td>
<td>Community</td>
<td>Clinic</td>
<td>Community</td>
</tr>
<tr>
<td>&lt;0.5%</td>
<td>90 (35.7%)</td>
<td>3 (3.2%)</td>
<td>216 (85.7%)</td>
<td>33 (34.8%)</td>
<td>164 (65.1%)</td>
<td>36 (37.9%)</td>
</tr>
<tr>
<td>~3%</td>
<td>115 (45.6%)</td>
<td>17 (17.9%)</td>
<td>20 (7.9%)</td>
<td>17 (17.9%)</td>
<td>63 (25.0%)</td>
<td>15 (15.8%)</td>
</tr>
<tr>
<td>~15%</td>
<td>30 (11.9%)</td>
<td>26 (27.4%)</td>
<td>14 (5.9%)</td>
<td>20 (21.1%)</td>
<td>14 (5.6%)</td>
<td>16 (16.8%)</td>
</tr>
<tr>
<td>~50%</td>
<td>13 (5.2%)</td>
<td>23 (24.2%)</td>
<td>2 (8.0%)</td>
<td>22 (23.2%)</td>
<td>7 (2.8%)</td>
<td>9 (9.5%)</td>
</tr>
<tr>
<td>70%+</td>
<td>4 (1.6%)</td>
<td>26 (27.4%)</td>
<td>0 (0.0%)</td>
<td>3 (3.2%)</td>
<td>4 (1.6%)</td>
<td>19 (20.0%)</td>
</tr>
</tbody>
</table>

Note. DAWBA= Development and Well-being Assessment.
Table 2: Bivariate correlations of DAWBA bands and corresponding disorders (expert diagnosis) in the community (n=252) and the clinic sample (N = 95)

<table>
<thead>
<tr>
<th></th>
<th>Community sample</th>
<th>Clinic sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any problem/disorders</td>
<td>.62***</td>
<td>.53***</td>
</tr>
<tr>
<td>Emotional problem/disorders</td>
<td>.31***</td>
<td>.67***</td>
</tr>
<tr>
<td>Behavioural problem/disorders</td>
<td>.59***</td>
<td>.60***</td>
</tr>
</tbody>
</table>

Note: *** = significance (two sided), p < .001.
Table 3: Bivariate correlations of SDQ parent and youth scales in the community (n=252) and the clinic sample (n = 95)

<table>
<thead>
<tr>
<th></th>
<th>Community sample</th>
<th>Clinic sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDQ total score</td>
<td>.46***</td>
<td>.20 n.s.</td>
</tr>
<tr>
<td>SDQ impact</td>
<td>.45***</td>
<td>.04 n.s.</td>
</tr>
<tr>
<td>SDQ emotion problems</td>
<td>.36***</td>
<td>.42***</td>
</tr>
<tr>
<td>SDQ behaviour problems</td>
<td>.38***</td>
<td>.37***</td>
</tr>
<tr>
<td>SDQ hyperactivity</td>
<td>.49***</td>
<td>.47***</td>
</tr>
</tbody>
</table>

Note: * = significance (two sided), p < .05, ** = significance (two sided), p < .01, *** = significance (two sided), p < .001.
Table 4: Ordinal regressions and Firth’s biased reduced logistic regressions with SDQ parent and youth measures as predictors of DAWBA bands/ disorders in the community sample (N = 252)

<table>
<thead>
<tr>
<th></th>
<th>Any problem/disorders</th>
<th>Emotional problem/disorders</th>
<th>Behavioural problem/disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate (95%CI)</td>
<td>OR (95%CI)</td>
<td>Estimate (95%CI)</td>
</tr>
<tr>
<td>SDQ total/impact score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent SDQ total score</td>
<td>0.67 (0.34-1.01)**</td>
<td>0.69 (0.11-1.27)*</td>
<td>0.31 (-0.15-0.78)</td>
</tr>
<tr>
<td>Parent SDQ impact</td>
<td>0.27 (-0.05-0.59)</td>
<td>0.12 (-0.33-0.57)</td>
<td>-0.25 (-0.73-0.23)</td>
</tr>
<tr>
<td>Youth SDQ total score</td>
<td>0.49 (0.19-0.78)**</td>
<td>0.54 (-0.04-1.14)</td>
<td>0.62 (0.18-1.06)**</td>
</tr>
<tr>
<td>Youth SDQ impact</td>
<td>0.62 (0.30-0.94)**</td>
<td>0.65 (0.21-1.16)**</td>
<td>0.45 (0.13-0.77)**</td>
</tr>
<tr>
<td>SDQ subscales</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent SDQ emotion problems</td>
<td>--</td>
<td></td>
<td>0.43 (0.10-0.76)*</td>
</tr>
<tr>
<td>Youth SDQ emotion problems</td>
<td>--</td>
<td></td>
<td>0.89 (0.49-1.30)**</td>
</tr>
<tr>
<td>SDQ subscales</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent SDQ behaviour problems</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Youth SDQ behaviour problems</td>
<td>--</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. SDQ= Strengths and Difficulties Questionnaire, DAWBA= Development and Well-being Assessment. OR=odds ratio. * = significance (two sided), p < .05, ** = significance (two sided), p < .01, *** = significance (two sided), p < .001. Age and male gender was included as covariates in the analyses.
Table 5: Ordinal and logistic regressions with SDQ parent and youth measures as predictors of DAWBA bands/disorders in the clinical sample (N = 95)

<table>
<thead>
<tr>
<th>SDQ total/impact score</th>
<th>Any problem/disorders</th>
<th>Emotional problem/disorders</th>
<th>Behavioural problem/disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate (95%CI)</td>
<td>OR (95%CI)</td>
<td>Estimate (95%CI)</td>
</tr>
<tr>
<td>Parent SDQ total score</td>
<td>1.02 (0.53-1.51)***</td>
<td>1.65 (0.89-3.07)</td>
<td>0.21 (-0.23-0.63)</td>
</tr>
<tr>
<td>Parent SDQ impact</td>
<td>0.19 (-0.25-0.62)</td>
<td>0.93 (0.51-1.67)</td>
<td>0.28 (-0.15-0.72)</td>
</tr>
<tr>
<td>Youth SDQ total score</td>
<td>0.50 (0.05-0.94)*</td>
<td>2.57 (1.32-5.01)**</td>
<td>0.42 (-0.01-0.85)</td>
</tr>
<tr>
<td>Youth SDQ impact</td>
<td>0.13 (-0.30-0.56)</td>
<td>1.17 (0.63-2.17)</td>
<td>0.54 (0.11-0.97)*</td>
</tr>
<tr>
<td>SDQ subscales</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent SDQ emotion problems</td>
<td>--</td>
<td>--</td>
<td>0.54 (0.10-0.97)*</td>
</tr>
<tr>
<td>Youth SDQ emotion problems</td>
<td>--</td>
<td>--</td>
<td>0.91 (0.44-1.38)**</td>
</tr>
<tr>
<td>SDQ subscales</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent SDQ behaviour problems</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Youth SDQ behaviour problems</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Note. SDQ= Strengths and Difficulties Questionnaire, DAWBA= Development and Well-being Assessment. OR=odds ratio. * = significance (two sided), p < .05, ** = significance (two sided), p < .01, *** = significance (two sided), p < .001. Age and male gender was included as covariates in the analyses.
**Figure 1.** Receiver Operating Characteristic analyses of the SDQ total and impact score to predict emotional disorders in male and female adolescents in the clinic sample (N = 95)

*Note.* SDQ = Strengths and Difficulties Questionnaire.
Male adolescents (n=63)

Female adolescents (n=32)
Click here to access/download
Supplementary Material
CAPMH_DATA_OCT2016.xlsx