

ABSTRACT

Objective: To examine screening efficiency for preschool psychopathology by comparing the Strengths and Difficulties Questionnaire against diagnostic information, and to determine the added value of impact scores and teacher information.

Method: Using a two-phase sampling design a population-based sample of 845 4-year-olds was recruited from community health check-ups in Trondheim, Norway; screen score stratified and oversampled for high screening scores. Blinded to screen ratings, DSM-IV diagnoses were assigned using the Preschool Age Psychiatric Assessment interview, against which the Strengths and Difficulties Questionnaire scores were compared through Receiver Operating Characteristics analysis.

Results: Emotional and behavioral disorders were identified through parent ratings with a specificity of 88.8% (87.0-90.6%) and a sensitivity of 65.1% (51.6-78.6%). The negative predictive value was 97.9% (96.8-98.9%), whereas the positive predictive value was 24.2% (18.0-30.3%) at a prevalence of 5.2%. Parental ratings identified more behavioral disorders (79.3%) than emotional disorders (59.2%). Screening for any disorder was somewhat less efficient; specificity 88.9% (87.0-90.7%), sensitivity 54.2% (41.8-66.6%), negative predictive value 96.4% (95.0-97.8%) and positive predictive value 25.9% (19.6-32.2%) at a prevalence of 6.7%. The AUC value was 0.83 (0.76-0.90) for emotional and behavioral disorders and 0.76 (0.68-0.83) for any disorder. The prediction accuracy was not improved by impact scores or teacher information.

Conclusions: The results indicate that preschoolers' emotional and behavioral disorders can be screened with the same efficiency as in older children and adults. Other disorders were identified to a lesser extent. Further research should explore the potential of preschool screening to improve early detection and subsequent intervention.

A large proportion of adult psychiatric disorders emerge early in life.^{1,2} Prospective studies indicate that later behavioral and emotional problems begin already in the preschool years.³⁻⁵ Available preschool studies report prevalences varying between 7% and 26%,⁶⁻¹¹ figures which are comparable to those found among school-aged children, adolescents and adults.¹² The potential long-term effects of treating preschool psychopathology may thus be vast. Unfortunately, the majority of children suffering from mental disorders do not receive treatment,^{1,13} which also appears to be the case for preschoolers.¹⁴

Because preschool children in need of treatment are seldom referred to child psychiatric services, we miss the opportunity to target the earliest time point when mental illness develops. Screening for mental health problems at the community level may thus be warranted. However, this depends on whether efficient screens for preschool disorders are available. Whereas the criteria for nearly all psychiatric disorders are similar throughout the lifespan, preschoolers' mental health problems are expected to be partly different from those of older children (e.g., more oppositional defiant disorders and separation anxiety, less depressive disorders and conduct disorders) and possibly also to have a different presentation.¹⁵ The efficiency of screens for preschoolers' mental health problems must therefore be evaluated specifically, downward extrapolation from results obtained for older children may not be accurate. However, an assessment of the efficiency of screening in a preschool community population has not yet been conducted. Thus, evaluating the screening efficiency of a commonly used screening instrument is the main goal of this study.

The Strengths and Difficulties Questionnaire (SDQ)¹⁶ has proven efficient in detecting psychopathology among older children.¹⁷⁻²¹ Moreover, its brevity, and the inclusion of competencies and strengths makes the SDQ more acceptable to respondents than other screening instruments. Parents and daycare providers are the only viable sources of information at the community level. However, we do not know whether information from both sources will increase accuracy. Additionally, the SDQ impact supplement provides an opportunity to assess whether diagnostic accuracy is improved by impairment ratings.

The effectiveness of a screening instrument varies by prevalence – it is much more difficult to accurately detect and particularly to avoid a high rate of false positives with uncommon disorders than more prevalent disorders. The prevalence in the present population¹¹ was in line with Scandinavian findings²²⁻²⁵ and fairly low compared to other preschool studies in the US.⁶⁻¹⁰ Thus, to increase the generalizability of our results, we will determine screening efficiencies for the most common range of prevalences.

By testing the SDQ against diagnostic information from a large community sample of 4 year olds, we aim to determine: 1) the overall screening efficiency for parent and teacher versions of the SDQ, 2) whether teachers add to the screening efficiency above parent information, 3) whether the impact score adds to the screening efficiency above the symptom score, 4) the optimal cut-point for SDQ and 5) the screening efficiency for various prevalences.

METHODS

Recruitment and participants

All parents and children attending the community health check-up for 4-year-olds in the city of Trondheim, Norway, were invited to participate. In total, 97.2% of invited families presented at the well child clinic, and 82.1% (2475) of the eligible families consented to participate. These parents filled out the SDQ P4-16 version.^{16,26} SDQ scores of school age children in Scandinavia^{22,25} are much lower than in many other countries, i.e. UK and the US. When stratifying, commonly used cut-offs were therefore replaced to reflect this by dividing the SDQ total difficulties score into four strata with the following cut offs: 0-4, 5-8, 9-11 and 12-40. Based on a random number generator, defined proportions of parents in each stratum were drawn to participate in a structured diagnostic interview concerning the child's mental health, amounting to a total of 1250. Because there was no pre-existing information concerning the distribution of SDQ scores among preschoolers at the time, we based the sampling on information from school aged Scandinavian children and heavily over-sampled those close to the upper 10th percentile (0.89), applied less oversampling among the moderately high scoring children (0.70), close to no oversampling among those supposed near the median (0.48) and under-sampled the low scoring children (0.37). Parents provided interview information on 80% (995) of the sample in this phase, of which 87% (863) had a postal questionnaire from teachers returned. The number of drop-outs at this stage was equally distributed across the four SDQ strata (24.5%, 18.7%, 23.7%, and 20.8%, respectively) and gender (girls: 22.9%, boys: 20.9%). The study was approved by the Regional Committee for Medical and Health Research Ethics.

The analyses reported in this paper are based on children with SDQ symptom and impact scores from both informants as well as interview information available (N=845). Descriptive information about the sample is provided in Table 1. The mean age of the children was 53.0 months (range 46.3-63.0, SD=2.1). All children were in governmental sponsored daycare centers. The population of Trondheim, where recruitment took place, is similar to the national average on several key indicators. The educational level of the parents in the sample was

virtually identical to the population's level, but significantly more were divorced (7.6%) than in the population (2.1%).¹¹

/ Table 1 near here/

Measures

Screening Scale

The parent and teacher versions of the SDQ (SDQP4-16 and SDQT4-16, respectively) are intended for children aged 4 to 16 years. Of the five 5-item subscales (emotional problems, conduct problems, hyperactivity, peer problems and pro-social behavior), the former four are summed to create a "total difficulties score" ranging from 0 to 40. The additional impact supplement captures the perceived difficulties, chronicity, impact and burden, where the impact score has shown to be the most discriminating predictor of disorders.²⁶ An impact score is computed adding scores on distress and social incapacity, giving ratings ranging between 0 and 10 for the parent questionnaire and between 0 and 6 for the teacher questionnaire. The impact scores significantly correlated with the PAPA total incapacity score; $r=0.42, p < .001$ (parent) and $r=0.26, p < .001$ (teacher).

A review reporting on 48 studies concluded that the psychometric properties of the SDQ for 4- to 12-year-olds were strong.²⁷ Two recent studies with younger age groups in the Netherlands (5- to 6-year-olds) and Denmark (5-year-olds) replicated the reliability and validity and confirmed the five-factor structure of the SDQ among preschoolers.^{28,29} The Norwegian version has been validated in several large studies.^{22,25} In our sample, Cronbach's alpha for the total difficulties score was 0.77 for the parent SDQ³⁰ and 0.86 for the teacher SDQ. The correlations between the parent and teacher SDQ ratings were $r=0.27, p < .001$ for the total difficulties score and $r=0.34, p < .001$ for the impact score.

Diagnostic Assessment

Psychiatric diagnoses were assigned based on the Preschool Age Psychiatric Assessment (PAPA), a semi-structured psychiatric interview for completion by parents of children aged 2 to 6 years.⁷ The interviewer-based structure requires the interviewer to ensure that the subjects understand the questions being asked and provide clear information on the behavior or feelings relevant to the symptom. Reported symptoms are held against pre-specified levels of severity and supplemented with the onset date and frequency of occurrence when relevant. Computerized algorithms implementing the Diagnostic and Statistical Manual of Mental Disorders (fourth edition)

(DSM-IV)³¹ were used to generate diagnoses. Problems regarding sleep onset and night walking were defined according to the Anders criteria³² and bipolar disorders according to modifications suggested by Luby and Belden.³³ Based on the World Health Organization's International Classification of Functioning, Disability and Health³⁴, impairment (disability) in 19 areas of functioning resulting from each group of symptoms was assessed. Thus, the presence of a symptom was followed by an evaluation of potential disability in three different settings (home, daycare or other settings). The PAPA has shown acceptable test-retest reliability.⁷

Interviewers (n=7) had at least a bachelor's degree in relevant fields and extensive prior experience in working with children and families. They were trained by the group who developed the measure and were blind to the SDQ results. To evaluate inter-rater reliability, 9% of the interview audio recordings were re-coded by blinded raters. Pairs of raters obtained the following multivariate inter-rater reliabilities;³⁵ ADHD: $k=.96$; ODD: $k=.89$; CD: $k=.78$; any anxiety disorder: $k=.89$; any depressive disorder: $k=.86$, any sleep disorder: $k=.87$; encopresis: $k=.92$; any disorder: $k=.83$.

Because SDQ taps into broad categories of problems instead of specific and rare disorders (e.g., tics), the PAPA diagnoses were joined together to create the following broad groupings of DSM-IV diagnoses; (1) a group of 'emotional disorders' consisting of major depression (MDD), dysthymia, depression not otherwise specified (NOS), separation anxiety disorder (SAD), generalized anxiety disorder (GAD), social phobia, specific phobia, agoraphobia, selective mutism and obsessive-compulsive disorder (OCD) ($k=.82$); (2) a group of 'behavioral disorders' consisting of attention-deficit/hyperactivity disorder (ADHD), oppositional defiant disorder (ODD) and conduct disorder (CD) ($k=.84$); (3) a group of 'emotional and behavioral disorders' comprising the disorders included in (1) and (2) ($k=.86$); and (4) a group of 'any disorder' consisting of disorders in (1) and (2) plus motor tics, vocal tics, trichotillomania, parasomnias and dyssomnias ($k=.83$). Given the high rate of encopresis in this age group, the diagnosis was excluded to prevent it from falsely influencing the scales estimated screening efficiency.³⁶

Statistical analysis

The screening efficiency of the parent and teacher scales was evaluated using Receiver Operating Characteristic Curve (ROC) analysis, which determines the area under the curve (AUC) for the scales against the diagnostic groups. The AUC expresses the probability that a randomly chosen subject with a disorder and a randomly chosen subject without a disorder would be correctly distinguished based on their screening scale scores. Hosmer and Lemeshow³⁷ provide general guidelines for interpreting AUC values as follows: AUC=0.5 (no

discrimination), $0.7 \leq \text{AUC} < 0.8$ (acceptable discrimination), $0.8 \leq \text{AUC} < 0.9$ (excellent discrimination) and $\text{AUC} \geq 0.9$ (outstanding discrimination). The potential added value of combining scales to increase the prediction of diagnosis was examined through logistic regression. By estimating the AUC of bivariate models, we determined whether including the teacher or impact scales significantly improved prediction accuracy above the regular parent total difficulties scale. Probability-weighted versions of the AUC and ROC, together with 95% confidence intervals (CI), were computed using Roger Newson's programs, `-somersd-` and `-senspec-`, which are available for download in Stata.^{38,39} Somersd computes the Harrell's C, an equivalent to the AUC,⁴⁰ referred to as the AUC here.

The sensitivity/specificity pairs generated through the ROC analysis were further used to select a threshold for identification of clinical cases. At a given cut-point sensitivity shows the proportion who receives a positive screen among diagnosed positives, whereas specificity denotes the proportion of children who receive a negative screen among diagnosed negatives. The positive predictive value (PPV) and negative predictive value (NPV) were calculated to provide a more comprehensive evaluation of the screening efficiency. The PPV provides the probability that a child who screens positive has the condition, whereas the NPV expresses the probability that a child who screens negative is a non-case. The sensitivity and specificity are more stable across populations than are the positive and negative predictive values.⁴⁰ Thus, the PPV and NPV of the SDQ in preschool populations with a different prevalence can be estimated from data on sensitivity and specificity as follows:

$$\text{PPV} = \frac{\text{Sensitivity} \times \text{Prevalence}}{\text{Sensitivity} \times \text{Prevalence} + (1 - \text{Specificity})(1 - \text{Prevalence})}$$

$$\text{NPV} = \frac{\text{Specificity} \times (1 - \text{Prevalence})}{(1 - \text{Sensitivity}) \times \text{Prevalence} + \text{Specificity} \times (1 - \text{Prevalence})}$$

Along with the prevalence found in the present sample, screening efficiency for prevalences of 10%, 15% and 20% were calculated.

The analyses were carried out using inverse probability weighting corresponding to the drawing probabilities in the four strata, and consequently, the results can be interpreted as estimates appropriate for 4-year-olds in the community population. All analyses were performed in Stata 11.⁴¹

RESULTS

Screening scale distribution

The parent total difficulties score presented with a mean of 5.6 (95% CI: 5.4-5.9), whereas the corresponding value for the teacher scale was 5.1 (95% CI: 4.8-5.4). The sample proportions for cut-points 8-14 of the parent total difficulties scale are displayed in Table 3. The 90th percentile was ≥ 11 whereas the 80th percentile was ≥ 9 .

Overall screening efficiency

As shown in Table 2, the parent total difficulties scale had excellent discrimination for emotional disorders, behavioral disorders and the combination of the two. For behavioral disorders alone, the AUC value approached outstanding discrimination. Acceptable discrimination was obtained for any disorder. The AUCs were considerably lower for the teacher total difficulties scale. Moreover, neither impact scale performed at the level of the parent total difficulties scale.

/ Table 2 near here/

Added value analyses

Because the parent SDQ total difficulties scale demonstrated excellent screening efficiency, it served as the starting scale for the added value analyses. Only slight increases in the AUCs were obtained by adding the teacher total difficulties scale or impact scores, none of which were significant (all $p > .05$). Taken together, these results suggest that in our preschool sample, only the parent SDQ total difficulties scale is needed to obtain maximum screening efficiency for the diagnostic groups of emotional/behavioral disorders, separately and combined, and any disorder.

Optimal cut-point for diagnosis

The cut-point maximizing the sum of sensitivity and specificity for the parent SDQ total difficulties scale was found at a score of 10 or more (Table 3). Raising the cut-point by one would lead to a considerable decrease in

sensitivity, whereas further lowering the cut-point would imply relatively larger decrease in specificity than increase in sensitivity. The scale performed quite stable ruling out a diagnosis (identifying non-cases as reflected by the specificity) across diagnostic categories. However, it was more variable ruling in a diagnosis (identifying true cases as reflected by the sensitivity), being less sensitive to emotional disorders than to behavioral disorders and less sensitive to the any disorder category compared to the emotional and behavioral disorders category.

/ Table 3 near here/

Screening efficiency for varying prevalences

Table 4 reports on the screening efficiency of the parent total difficulties scale in terms of sensitivity, specificity, PPV and NPV at a cut-point of 10. The prevalences in the present sample were 3.3% (CI: 2.3-4.4%) for emotional disorders, 3.0% (CI: 2.1-4.0%) for behavioral disorders, 5.2% (CI: 3.9-6.5%) for emotional and behavioral disorders and 6.7% (CI: 5.1-8.2%) for any disorder. Table 4 further illustrates how the PPV and NPV vary as a function of prevalence. In a population where emotional and behavioral disorders sum up to a prevalence of 10%, the parent total difficulties scale would obtain a PPV of 39.2%. In other words; doubling the prevalence from 5 to 10, results in a 15% increase in the PPV. The NPV values are more stable across prevalence; the same 5% increase in prevalence would result in a 2% decrease in the NPV (from 97.9 to 95.8) for the parent total difficulties scale.

/ Table 4 near here/

DISCUSSION

Drawing on diagnostic interview data from a large community sample, we assessed whether common psychiatric diagnoses can be efficiently screened for in preschool-aged children. The results showed that the parent-rated SDQ4-16 was able to detect preschoolers with emotional and behavioral disorders with reasonable efficiency. The impact scores did not add to the screening efficiency beyond the symptom scores, and the teacher ratings did not increase diagnostic accuracy above the parent ratings.

Overall screening efficiency

The capacity to discriminate cases from non-cases in this study (AUC=0.83) is on par with results from earlier studies of school-aged children using the SDQ²⁷ and of adults using the K6 and K10.^{42,43} The estimated specificity and negative predictive value were high (approaching and above 90%), meaning that the SDQ

identified non-cases with great accuracy. However, sensitivity for all diagnostic categories was considerably lower, ranging from 54-79% at the selected cut-point of 10. When examining parent predictions based on a 90th percentile cut-point in school-aged community children, Goodman²¹ found a sensitivity of 47% for any DSM-IV diagnosis. When applying a scoring algorithm for parent SDQs, the sensitivities for community children aged 5-10 years ranged from 29.8% for any psychiatric disorder to 53.9% for depressive disorders.¹⁹ Moreover, when evaluating the K6 scale as a screen for severe mental illness (SMI) among adults, Kessler et al.⁴² reported a sensitivity of 36% at the optimal cut-point. This finding indicates that we can screen for psychopathology in preschool-aged children with comparable efficiency as in older children and adults. Moreover, the same tendency can be seen across all age groups; when screening in community samples, the proportion of true negatives (NPV) is high, but the proportion of true positives (PPV) is substantially lower. In screening tests, where the first priority is to reduce the rate of false negatives, this sort of over-inclusiveness may be considered acceptable. The consequent increased rate of false positive cases may lead to unnecessary costs in psychiatric follow-up evaluations, as well as stress and perhaps worry among those who are falsely screened as positive. However, these costs must be weighed against the financial and human burden of failing to identify and treat psychiatric disorders that would have gone undetected if the screening program was not undertaken.

It should be noted that the reported PPVs and NPVs are affected by the lower prevalences in the present sample. As prevalence increases, the PPV increases, while the NPV decreases. In many non-Nordic preschool populations, where the prevalence of psychiatric disorders is expected to be higher, a higher rate of true positives would be detected (increased PPV), but a somewhat larger proportion would be false negatives (decreased NPV).

For the any disorder group, screening efficiency was acceptable (AUC=0.76). However, a sensitivity of 54% means that almost half of the cases in the present sample were not identified by the SDQ. The same tendency was observed in a British community sample of 5- to 10-year-olds; the sensitivity of the parent-rated SDQ was poorer for any psychiatric disorder than for emotional and behavioral disorders.¹⁹ In addition to emotional and behavioral disorders, the any disorder group in the present sample consisted of disorders that are not considered by any of the 20 SDQ problem score items (e.g., tics and sleep problems). Instead, these 20 items tap into emotional and behavioral aspects of child functioning, which may explain why the screening efficiency drops when screening disorders outside of this realm. Goodman and Scott²⁰ concluded that the restricted amount of problems covered in the SDQ makes it unsuitable for studies or clinical assessments that require coverage of a broad range of childhood psychopathology. To uncover all of the specific psychopathologies, a range of items would need to be added to the scale, which would run counter to a key aspect of a screening measure; brevity.

For screening purposes, a shorter scale that accurately predicts the common disorders would arguably be preferred over a longer, less user-friendly scale. Our results suggest that the SDQ is a viable option for screening in a community setting, given that organizational systems to support the management, treatment and follow-up of those screened positive are in place.

Informant and disorder specific variations

The screening efficiency of the SDQ4-16 depends on the informant. The lower performance of teacher ratings for all diagnostic categories disagrees with findings obtained for school-aged children, for whom teachers provide information of roughly equal predictive value as parents.²⁷ As put forth by Elberling et al.,⁴⁴ this observation may be explained by the difference in observer context for preschool teachers versus school teachers. In the present sample of 4 year olds, all children were attending a daycare center with a less demanding environment than school, with fewer rules and less structure. In non-structured activities, attention problems, hyperactivity and conduct problems may be concealed or more easily interpreted as merely normal variations of preschoolers' inattentiveness, activity levels or impulsive aggressiveness.

The parents in our sample detected more externalizing/behavioral disorders than internalizing/emotional disorders. Such lower detection of internalizing disorders has been seen in other studies^{45,46} and, as noted by others,^{47,48} it underpins the uncertainty in solely relying on parent-report. Moreover, the observation could be explained by reporting bias, described by Heiervang et al.⁴⁶ as Norwegian adults taking a more 'normalizing' view of emotional symptoms when filling out screening questionnaires. However, if this is the case, our sample is more likely to have attenuated the screening efficiency of the SDQ rather than exaggerating it. That is, the sensitivity for emotional disorders may be higher in another population.

Neither the teacher nor parent impact scores added to the screening efficiency beyond the parent symptom score in the present study. For emotional disorders, this finding may be due to the aforementioned 'normalizing' view of Norwegian adults. Greater adult permissiveness and lower expectations towards preschoolers could also explain the poorer discriminative value than observed in older children²⁶ and possibly make the impact score less relevant for screening purposes in preschool populations. Additionally the impact score may not be developmentally appropriate for preschool age children.

Methodological considerations

As the reviews indicate that the reliability and validity of the SDQ in western countries are comparable²⁷ and assuming that sensitivity and specificity are reasonably stable across populations⁴⁰, we expect that our results for screening efficiency can be generalized to other preschool populations. The estimated cut-point in the present sample is comparable to findings of younger children (4-7 years) in Northern Europe^{28,29} but lower than American⁴⁹ and British¹⁶ findings, which speaks to the importance of deriving cut-points from the population of interest. It should be underscored that the choice of cut-point affects the estimated screening efficiency. In community samples, where milder symptomatology predominates, true cases are more likely to be missed than in clinical samples. Thus, a low cut-point may be justified to increase screening sensitivity, avoiding false negatives. In the present sample a cut-point of 10 identifies most of the non-cases (high specificity) and keeps the rate of false negatives low (high NPV). On the other hand, the sensitivity rates are moderate and a considerable portion is falsely identified as disordered (low PPV). However, further lowering the cut-point to increase sensitivity would cause only a modest gain in detection of true cases and a relatively larger increase in false positives, which was not considered a better trade-off. Choosing a cut-point that maximizes sensitivity in a community low-risk sample yields low PPV and thus a larger proportion of false positives. This must be balanced to avoid unnecessary burden on the health services and on children and parents who are not in need of treatment. More complex calculations taking into consideration the cost of assessment and/or treatment might however militate in favor of adjusting the cut-point.

The findings should be viewed in the light of some methodological limitations. Our subjects were mostly of Norwegian origin; thus, the findings may not generalize to populations of more heterogeneous composition. The relatively small number of uncommon disorders (n=12) may have limited our ability to evaluate the utility of the SDQ to screen for these disorders. It should also be noted that the parent-reported SDQ scores were compared with the PAPA interview, which also was derived from parental information. Although the PAPA interview clearly is interviewer-based, it would have strengthened the study if additional comparisons (e.g. clinician rating, blind diagnosis of parent-child interaction) also served as a criterion standard. Additionally, the SDQ3-4 may have performed differently in our sample of 4-year-olds than the SDQ4-16 that we used; however, the SDQ4-16 was chosen to be able to longitudinally track change and stability in the SDQ scores.

REFERENCES

1. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005;62(6):593-602.
2. Kim-Cohen J, Caspi A, Moffitt TE, Harrington H, Milne BJ, Poulton R. Prior juvenile diagnoses in adults with mental disorder: developmental follow-back of a prospective-longitudinal cohort. *Arch Gen Psychiatry*. Jul 2003;60(7):709-717.
3. Stevenson J, Goodman R. Association between behaviour at age 3 years and adult criminality. *Br J Psychiatry*. Sep 2001;179:197-202.
4. Mesman J, Koot HM. Early preschool predictors of preadolescent internalizing and externalizing DSM-IV diagnoses. *J Am Acad Child Psy*. Sep 2001;40(9):1029-1036.
5. Caspi A, Moffitt TE, Newman DL, Silva PA. Behavioral observations at age 3 years predict adult psychiatric disorders. Longitudinal evidence from a birth cohort. *Arch Gen Psychiatry*. Nov 1996;53(11):1033-1039.
6. Earls F. Application of DSM-III in an epidemiological study of preschool children. *Am J Psychiatry*. Feb 1982;139(2):242-243.
7. Egger HL, Erkanli A, Keeler G, Potts E, Walter BK, Angold A. Test-Retest Reliability of the Preschool Age Psychiatric Assessment (PAPA). *J Am Acad Child Psy*. May 2006;45(5):538-549.
8. Keenan K, Shaw DS, Walsh B, Delliquadri E, Giovannelli J. DSM-III-R Disorders in Preschool Children From Low-Income Families. *Journal of the American Academy of Child & Adolescent Psychiatry*. 1997;36(5):620-627.
9. Lavigne JV, Gibbons RD, Christoffel KK, et al. Prevalence rates and correlates of psychiatric disorders among preschool children. *J Am Acad Child Psy*. Feb 1996;35(2):204-214.
10. Lavigne JV, Lebailly SA, Hopkins J, Gouze KR, Binns HJ. The prevalence of ADHD, ODD, depression, and anxiety in a community sample of 4-year-olds. *J Clin Child Adolesc Psychol*. May 2009;38(3):315-328.
11. Wichstrom L, Berg-Nielsen TS, Angold A, Egger HL, Solheim E, Sveen TH. Prevalence of psychiatric disorders in preschoolers. *Journal of child psychology and psychiatry, and allied disciplines*. Jun 2012;53(6):695-705.
12. Egger HL, Angold A. Common emotional and behavioral disorders in preschool children: presentation, nosology, and epidemiology. *Journal of Child Psychology and Psychiatry*. Mar-Apr 2006;47(3-4):313-337.
13. Kessler RC, Demler O, Frank RG, et al. Prevalence and treatment of mental disorders, 1990 to 2003. *N Engl J Med*. Jun 16 2005;352(24):2515-2523.
14. Lavigne JV, Arend R, Rosenbaum D, Binns HJ, Christoffel KK, Gibbons RD. Psychiatric disorders with onset in the preschool years: I. Stability of diagnoses. *J Am Acad Child Psy*. Dec 1998;37(12):1246-1254.
15. Angold A, Egger HL. Preschool psychopathology: lessons for the lifespan. *Journal of child psychology and psychiatry, and allied disciplines*. Oct 2007;48(10):961-966.
16. Goodman R. The strengths and difficulties questionnaire: A research note. *J Child Psychol Psys*. Jul 1997;38(5):581-586.
17. Malmberg M, Rydell AM, Smedje H. Validity of the Swedish version of the Strengths and Difficulties Questionnaire (SDQ-Swe). *Nord J Psychiat*. 2003;57(5):357-363.
18. Klasen H, Woerner W, Wolke D, et al. Comparing the German versions of the Strengths and Difficulties Questionnaire (SDQ-Deu) and the Child Behavior Checklist. *Eur Child Adolesc Psychiatry*. Dec 2000;9(4):271-276.
19. Goodman R, Ford T, Simmons H, Gatward R, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *Br J Psychiatry*. Dec 2000;177:534-539.
20. Goodman R, Scott S. Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: is small beautiful? *J Abnorm Child Psych*. Feb 1999;27(1):17-24.

21. Goodman R. Psychometric properties of the strengths and difficulties questionnaire. *J Am Acad Child Psy.* Nov 2001;40(11):1337-1345.
22. Obel C, Heiervang E, Rodriguez A, et al. The Strengths and Difficulties Questionnaire in the Nordic countries. *Eur Child Adolesc Psychiatry.* 2004;13 Suppl 2:II32-39.
23. Kristensen S, Henriksen TB, Bilenberg N. The Child Behavior Checklist for Ages 1.5-5 (CBCL/11/2-5): Assessment and analysis of parent- and caregiver-reported problems in a population-based sample of Danish preschool children. *Nord J Psychiat.* May 2010;64(3):203-209.
24. Rescorla L, Achenbach T, Ivanova MY, et al. Behavioral and emotional problems reported by parents of children ages 6 to 16 in 31 societies. *J Emot Behav Disord.* Fal 2007;15(3):130-142.
25. Heiervang E, Stormark KM, Lundervold AJ, et al. Psychiatric disorders in Norwegian 8- to 10-year-olds: an epidemiological survey of prevalence, risk factors, and service use. *J Am Acad Child Adolesc Psychiatry.* Apr 2007;46(4):438-447.
26. Goodman R. The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. *Journal of child psychology and psychiatry, and allied disciplines.* Jul 1999;40(5):791-799.
27. Stone LL, Otten R, Engels RC, Vermulst AA, Janssens JM. Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-year-olds: a review. *Clin Child Fam Psychol Rev.* Sep 2010;13(3):254-274.
28. Mieloo C, Raat H, van Oort F, et al. Validity and reliability of the strengths and difficulties questionnaire in 5-6 year olds: differences by gender or by parental education? *PLoS One.* 2012;7(5):e36805.
29. Niclasen J, Teasdale TW, Andersen AM, Skovgaard AM, Elberling H, Obel C. Psychometric properties of the Danish Strength and Difficulties Questionnaire: the SDQ assessed for more than 70,000 raters in four different cohorts. *PLoS One.* 2012;7(2):e32025.
30. Berg-Nielsen TS, Solheim E, Belsky J, Wichstrom L. Preschoolers' psychosocial problems: in the eyes of the beholder? Adding teacher characteristics as determinants of discrepant parent-teacher reports. *Child psychiatry and human development.* Jun 2012;43(3):393-413.
31. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (fourth edition).* Washington DC: American Psychiatric Association; 1994.
32. Anders T, Eiben L. Sleep disorders. In: C. Z, ed. *Handbook of infant mental health.* New York: The Guildford Press; 2000:326-338.
33. Luby JL, Belden A. Defining and validating bipolar disorder in the preschool period. *Dev. Psychopathol.* Fal 2006;18(4):971-988.
34. World Health Organization. *ICF: International Classification of Functioning, Disability and health.* Geneva, Switzerland: World Health Organization; 2001.
35. Janson H, Olsson U. A measure of agreement for interval or nominal multivariate observations by different sets of judges. *Educ. Psychol. Meas.* Feb 2004;64(1):62-70.
36. Carol Joinson P, Jon Heron P, Ursula Butler M, Alexander von Gontard M. Psychological Differences Between Children With and Without Soiling Problems. *Pediatrics.* 2006;117(5):10.
37. Hosmer DW, Lemeshow S. *Applied logistic regression.* 2nd ed. New York: Wiley; 2000.
38. Newson R. Confidence intervals for rank statistics: Somers' D and extensions. *The Stata Journal* 2006;6(3):309-334.
39. Newson R. senspec: Stata module to compute sensitivity and specificity results saved in generated variables <http://ideas.repec.org/c/boc/bocode/s439801.html>. Accessed 15.10.12, 2012.
40. Zhou XH, Obuchowski, N. A., McClish, D. K. *Statistical methods in diagnostic medicine.* 2 ed: Hoboken: John Wiley & Sons Inc.; 2011.
41. *Stata Statistical Software: Release 11* [computer program]. College Station, TX: StataCorp LP.2009.

42. Kessler RC, Barker PR, Colpe LJ, et al. Screening for serious mental illness in the general population. *Arch Gen Psychiatry*. Feb 2003;60(2):184-189.
43. Kessler RC, Green JG, Gruber MJ, et al. Screening for serious mental illness in the general population with the K6 screening scale: results from the WHO World Mental Health (WMH) survey initiative. *Int J Methods Psychiatr Res*. Jun 2010;19 Suppl 1:4-22.
44. Elberling H, Linneberg A, Olsen EM, Goodman R, Skovgaard AM. The prevalence of SDQ-measured mental health problems at age 5-7 years and identification of predictors from birth to preschool age in a Danish birth cohort: the Copenhagen Child Cohort 2000. *Eur Child Adolesc Psychiatry*. Sep 2010;19(9):725-735.
45. Wu P, Hoven CW, Bird HR, et al. Depressive and disruptive disorders and mental health service utilization in children and adolescents. *J Am Acad Child Psy*. Sep 1999;38(9):1081-1090; discussion 1090-1082.
46. Heiervang E, Goodman A, Goodman R. The Nordic advantage in child mental health: separating health differences from reporting style in a cross-cultural comparison of psychopathology. *Journal of child psychology and psychiatry, and allied disciplines*. Jun 2008;49(6):678-685.
47. Luby JL, Belden A, Sullivan J, Spitznagel E. Preschoolers' contribution to their diagnosis of depression and anxiety: uses and limitations of young child self-report of symptoms. *Child psychiatry and human development*. Dec 2007;38(4):321-338.
48. Kemper TS, Gerhardstein R, Repper KK, Kistner JA. Mother-child agreement on reports of internalizing symptoms among children referred for evaluation of ADHD. *J Psychopathol Behav*. Dec 2003;25(4):239-250.
49. Bourdon KH, Goodman R, Rae DS, Simpson G, Koretz DS. The Strengths and Difficulties Questionnaire: U.S. normative data and psychometric properties. *J Am Acad Child Psy*. Jun 2005;44(6):557-564.

Table 1. Sample characteristics

Characteristic		%
Gender of child	Male	49.2
	Female	50.8
Gender of parent informant	Male	15.7
	Female	84.3
Ethnic origin of biological mother	Norwegian	92.7
	Western countries	3.1
	Other countries	4.2
Ethnic origin of biological father	Norwegian	90.7
	Western countries	6.1
	Other countries	3.2
Biological parents' marital status	Married	55.6
	Cohabiting >6 months	32.7
	Separated	1.4
	Divorced	7.5
	Widowed	0.4
	Cohabited <6 months	1.2
	Never lived together	1.2
Informant parent's socioeconomic status	Leader	5.1
	Professional, higher level	26.6
	Professional, lower level	39.1
	Formally skilled worker	25.3
	Farmer/fisherman	0.6
	Unskilled worker	3.3
Parent's completed education	Did not complete junior high school	0
	Junior high school (10 th grade)	0.7
	Some education after junior high	5.9

	school	
	Senior high school (13 th grade)	16.6
	Some education after senior high school	3.2
	Some college or university education	7.2
	Bachelor degree	6.2
	College degree (3-4 years of study)	34.0
	Master degree or similar	21.9
	PhD completed or ongoing	4.3
Gross annual household income	0 – 225' NOK (0 – 40' USD)	3.0
	225 – 525' NOK (40 – 94' USD)	18.0
	525 - 900' NOK (94 – 161' USD)	51.7
	900' + NOK (161' + USD)	27.3

Note: NOK = Norwegian Kroner; USD = United States Dollars

Table 2. Overall screening efficiency, AUC

	Emotional disorders (n=39)	Behavioral disorders (n=40)	Emotional and behavioral disorders (n=63)	Any disorder (n=75)
	AUC (95% CI)	AUC (95% CI)	AUC (95% CI)	AUC (95% CI)
SDQ parent scales				
<i>Total difficulties</i>	0.804 (0.723-0.886)	0.890 (0.808-0.972)	0.828 (0.759-0.898)	0.755 (0.677-0.833)
<i>Impact</i>	0.639 (0.565-0.713)	0.708 (0.628-0.789)	0.657 (0.597-0.717)	0.616 (0.566-0.666)
SDQ teacher scales				
<i>Total difficulties</i>	0.702 (0.611-0.794)	0.755 (0.676-0.835)	0.695 (0.623-0.766)	0.645 (0.578-0.712)
<i>Impact</i>	0.650 (0.568-0.731)	0.673 (0.591-0.754)	0.641 (0.577-0.704)	0.603 (0.542-0.665)

Note: AUC = area under the curve; CI = confidence interval; SDQ = Strengths and Difficulties Questionnaire

Table 3. Receiver Operating Characteristics Analyses for the SDQ P4-16

SDQP Score	Sample proportion, %	Emotional disorders (n=39)		Behavioral disorders (n=40)		Emotional & behavioral disorders (n=63)		Any disorder (n=75)	
		Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec
8	6.5	0.669	0.765	0.907	0.771	0.734	0.777	0.607	0.776
9	4.5	0.618	0.830	0.822	0.835	0.667	0.841	0.555	0.841
10	4.0	0.592	0.875	0.793	0.880	0.651	0.888	0.542	0.889
11	2.5	0.410	0.910	0.650	0.917	0.485	0.921	0.400	0.921
12	2.8	0.359	0.935	0.593	0.941	0.435	0.945	0.348	0.944
13	1.0	0.339	0.963	0.483	0.966	0.358	0.970	0.289	0.970
14	1.4	0.299	0.972	0.440	0.975	0.307	0.978	0.249	0.978

Note: SDQ = Strengths and Difficulties Questionnaire, Sens = Sensitivity, Spec = Specificity

Table 4. Screening efficiency by prevalence

Diagnostic group and SDQ scale	Sensitivity (95% CI)	Specificity (95% CI)	Prevalence (%)	PPV (95% CI)	NPV (95% CI)
Emotional disorders					
(n=39)					
<i>SDQ parent total difficulties score (≥10)</i>	0.592 (0.417-0.766)	0.875 (0.856-0.895)	3.3	0.141 (0.091-0.191)	0.984 (0.975-0.993)
			10.0	0.345	0.951
			15.0	0.455	0.924
			20.0	0.542	0.896
Behavioral disorders					
(n=40)					
<i>SDQ parent total difficulties score (≥10)</i>	0.793 (0.639-0.947)	0.880 (0.862-0.899)	3.0	0.172 (0.118-0.225)	0.993 (0.987-0.999)
			10.0	0.423	0.975
			15.0	0.538	0.960
			20.0	0.623	0.944
Emotional and behavioral disorders					
(n=63)					
<i>SDQ Parent total difficulties score (≥10)</i>	0.651 (0.516-0.786)	0.888 (0.870-0.906)	5.2	0.242 (0.180-0.303)	0.979 (0.968-0.989)

10.0	0.392	0.958
15.0	0.506	0.935
20.0	0.592	0.911

Any disorder (n=75)

<i>SDQ parent total difficulties score (≥10)</i>	0.542 (0.418-0.666)	0.889 (0.870-0.907)	6.7	0.259 (0.196-0.322)	0.964 (0.950-0.978)
			10.0	0.352	0.946
			15.0	0.463	0.917
			20.0	0.550	0.886

Note: CI = confidence interval; NPV = negative predictive value; PPV = positive predictive value; SDQ = Strengths and Difficulties Questionnaire