CHAPTER 04

HOW RELIABLE AND VALID IS THE TEACHER VERSION OF THE STRENGTHS AND DIFFICULTIES QUESTIONNAIRE IN PRIMARY SCHOOL CHILDREN?


Submitted
ABSTRACT

Introduction: Reporting on emotional and behavioral problems using the Strengths and Difficulties Questionnaire (SDQ) is validated for parents, but not yet for teachers in a broad age range of children. We conducted a cross-sectional study in 4-10 years old school children to investigate if the SDQ-T can be used instead of the validated but lengthy Teacher’s Report Form (TRF) to acquire information about emotional and behavioral problems in the school community.

Methods: Teachers of 453 children from primary schools were approached. Teachers of 394 children (response rate 86.9%) with a mean age of 7.1 years filled in the SDQ-T (n = 387), the TRF (n = 349) or both (n = 342). We assessed reliability by calculating internal consistency and concurrent validity (using correlation coefficients, sensitivity, specificity) of the SDQ-T compared with the TRF.

Results: Internal consistency of the SDQ-T Total Difficulties Score (SDQ-T TDS; Cronbach α=0.80), hyperactivity/ inattention- (α = 0.86) and prosocial behavior (α = 0.81) was very good. Concurrent validity demonstrated a strong correlation of all subscales of the SDQ-T with the corresponding scale on the TRF (range 0.54-0.73), except for peer problems (0.46). Using a SDQ-T TDS cut-off score > 14, the SDQ-T had a good sensitivity (90%) and specificity (94%).

Discussion: The good reliability, validity and brevity of the SDQ-T make it an easily applicable questionnaire for obtaining information about emotional and behavioral problems from teachers in primary school children.
How reliable and valid is the teacher version of the SDQ | Chapter 4

INTRODUCTION

Mental health problems affect 10-20% of children and adolescents worldwide, and may strongly hamper their functioning. Mental health disorders are currently the leading causes of disability in children. Accumulating evidence shows that young children with mental health problems are progressively at risk for a range of negative outcomes. In adulthood these include psychiatric disorders, poor academic achievement and a lower socio-economic status. Prevention of these poor outcomes might be possible with early recognition and prompt mental health treatment.

Reliable and valid questionnaires to report mental health problems are available for young children. The parent version of the Strengths and Difficulties Questionnaire (SDQ) has quickly become one of the most utilized screening instruments because of its brevity and its ability to measure both problem behavior and competencies. The SDQ includes only 25 items, is freely available and translated in many languages. Recently, the parent version of the SDQ identified significant mental health concerns in a large sample of children and adolescents screened in pediatric primary care offices.

For clinicians it is important to have multiple informants report on emotional and behavioral problems in children because these problems might be highly situational. For example, children can be very hyperactive in the school environment, because of the interaction with their classmates, but not display this behavior in the home environment in a one-on-one situation. Furthermore, for most DSM-V mental health diagnosis in children both parental and teacher information is required, as symptoms must be present in two or more settings (e.g. at home and school) but also there must be clear evidence that the symptoms interfere with or reduce the quality of social or school functioning. Identifying the specific context in which children display emotional and behavioral problems and its impact on school functioning, may also facilitate treatment of these problems. Teachers are well acknowledged informants of children’s behavior because they see children on a daily basis in the school environment and have the opportunity to compare behavior of children of similar age every day.

A reliable and valid questionnaire, the Teacher’s Report Form (TRF), is available for teachers to assess the extent of a child’s emotional and behavioral problems. The TRF is the teacher version of the Achenbach Child Behavior Checklist (CBCL), with the same questions worded differently for teachers. Both questionnaires are often regarded as the gold standard among broadband behavior rating scales. For routine use, a major disadvantage of the TRF is its length, as it includes 113 items. Time and administrative burdens are reported to be important barriers for identifying mental health problems in the school environment. The use of a short, inexpensive, easy accessible questionnaire could facilitate obtaining information from teachers.
The parent- and adolescent-report versions of the SDQ have been shown to have good psychometric properties.\textsuperscript{6,10,16-20} In contrast, the validity of the teacher version of the SDQ (SDQ-T) has not been investigated in a broad age range of school children.\textsuperscript{10,21} Although a recent review reported the strong psychometric properties of the SDQ-T, in most studies the SDQ-T results have not been compared with a gold standard for teachers, but with a gold standard for parents (e.g. Child Behavior Checklist or the Conner’s Parent Symptom Questionnaire).\textsuperscript{16,19,22} Studies that did compare the SDQ-T with the TRF have some limitations; they studied a specific age range (children age 5-6 years only) and another study used the TRF-reference data of another population than the one studied.\textsuperscript{23-25}

Clinicians are increasingly held accountable to assess mental health problems in their practice and are mandated to collaborate with parents and teachers on these problems.\textsuperscript{26} It would be ideal for them to use the same, short, questionnaire for both parents and teachers to express their mental health concerns. The aim of our study was therefore to examine the reliability and validity of the SDQ-T with the TRF as a gold standard, in 4-10 years old children in primary school children.

**METHODS**

**Participants and Procedure**

We obtained data using a two-step procedure. In the first step directors of all 70 regular primary schools in the middle and eastern regions of the Netherlands were contacted and asked if they were interested to participate in the study. Seventeen schools consented to have teachers participate in the study (24.3%). Reasons for schools not to participate were time constraints or participation in other studies.

In the second step, information about the study was sent to all parents, and 1664 parents signed informed consent (response rate 40.3%). Since the TRF is a long questionnaire and time consuming for teachers to fill in, only 5 children per class were randomly chosen to participate in our study. This led to the inclusion of teachers and parents of 453 children. Of these, 22 (4.9%) of the children were lost to follow-up and 37 (8.2) had incomplete data.

We obtained data on child and family background characteristics; i.e. age, gender, ethnic background (country of origin of parents) and educational level (highest completed grade) of parents. Teachers of 394 children (response 86.9%) filled in the SDQ-T (n = 387), the TRF (n= 349) or both questionnaires (n =342). Figure 1 describes the data collection process. This study was approved by the ethics committee of the Radboud University Nijmegen, in the Netherlands.
Figure 1 Flow chart for data collection
MEASUREMENTS

Strengths and Difficulties Questionnaire
The SDQ consists of 25 items subdivided into four difficulties scales, emotional symptoms, conduct problems, inattention-hyperactivity, peer problems, and a separate fifth strength scale, prosocial behavior. All subscales had five questions each. An impact supplement inquires further about the existence, chronicity, and distress of problems, social- and learning impairment, and burden to others, these items can be summed to generate an “impact score”. Each item has to be scored on a 3-point scale with 0 = ‘not true’, 1 = ‘somewhat true’ and 2 = ‘certainly true’. An example of a question in the inattention-hyperactivity subscale is: “Restless, overactive, cannot stay still for long”. We used the Dutch teacher version of the SDQ 4-16 years including the impact supplement. The SDQ-T Total Difficulties Score (SDQ-T TDS) was calculated by aggregating the scores for the emotional symptoms, conduct problems, hyperactivity-inattention, and peer problems subscales (range 0-40). We dichotomized the SDQ-T TDS according to Goodman’s recommendation to define the highest 10% in a clinical range. In the present study population the clinical cut-off score for the SDQ-T TDS was found to be 14.

Teacher’s Report Form
The TRF has 113 items across eight syndrome scales (Anxious/Depressed, Withdrawn/Depressed, Somatic Complaints, Social Problems, Thought Problems, Attention Problems, Rule-breaking Behavior and Aggressive Behavior) and two broadband scales (internalizing and externalizing problems). Teachers have to indicate on 3-point scales the extent to which each item applies: 0 = ‘not’, 1 = ‘sometimes’, or 2 = ‘often’. An example question of the Attention Problems scale is “Can’t concentrate, can’t pay attention for long”. We used the validated Dutch version of the TRF. We computed the scores on the syndrome- and broadband scales, and the total problems scale. The total problem scale (TRF TPS) was dichotomized according to the Dutch cut-off score for the subclinical and clinical range.

ANALYSIS
All analyses were performed with IBM SPSS 20. First we described the background characteristics of our study population. Second, we examined the reliability, i.e. the extent to which items produce similar scores, of the SDQ-T. We therefore examined the scale structure of the SDQ-T by calculating internal consistencies. Standardized Cronbach’s alpha coefficients (α) were computed for the SDQ scales (emotional symptoms, conduct problems, hyperactivity / inattention, peer problems, prosocial behavior) impact score and total difficulties score. Cronbach’s α = 0.70 and below are generally considered as low, values between α = 0.70 and α = 0.80 as acceptable, and values above α = 0.80 as good.
Third, we assessed the concurrent validity, by determining the degree to which SDQ-T outcomes concurred with those on the TRF. We expected correlations between the SDQ scales and the TRF scales that rated similar problems. The emotional subscale of the SDQ-T was expected to correlate with the Internalizing broadband scale of the TRF. The conduct problem scale of the SDQ-T was expected to correlate with the Aggressive Behavior and Externalizing broadband scale of the TRF. The hyperactivity / inattention subscale of the SDQ-T was expected to correlate with both the Attention Problem and Externalizing scale of the TRF. The peer problem subscale of the SDQ-T was expected to correlate with the Social Problem scale of the TRF. The SDQ-T prosocial subscale, which includes questions about strengths, was thought to have a negative correlation with all problem scales of the TRF. Because of the non-normal distribution of both the SDQ-T and TRF concurrent validity was assessed using Spearman’s correlation of the SDQ-T with the TRF. Correlations below 0.30 are considered as small, correlations between 0.30 and 0.50 as medium, and correlations above 0.50 as strong.

We computed sensitivity (the proportion of children with emotional and behavioral problems who are correctly identified with the SDQ-T), specificity (the proportion of children without emotional and behavioral problems who are correctly identified without problems with the SDQ-T) and positive predictive value (the probability that children with a clinical SDQ-TDS truly have emotional and behavioral problems) with the TRF-TPS used as criterion. We further calculated the Area Under the Receiver Operator characteristic Curve (AUC) that integrates sensitivity and specificity across the various cut-offs of the dichotomized SDQ-T TDS. A value of the AUC of 1 reflects perfect accuracy of the SDQ-T to discriminate between children with- and without emotional and behavioral problems, and a value of 0.5 the absence of capacity to discriminate.

RESULTS

Background characteristics
The characteristics of the study population are presented in Table 1. The study population consisted of 202 boys (51.3%) and 192 girls (48.7%), with a mean age of 7.1 years (range 4.0 – 10.8 years). Most parents had medium (secondary vocational education) or higher educational (university or higher vocational education) level (Table 1). The SDQ-T mean scores and the clinical cut-off scores (p90) scores in the total sample and stratified by gender and age are presented in Table 2.
Chapter 4

Table 1 Social and demographic background of study population

<table>
<thead>
<tr>
<th></th>
<th>Number (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Boys</td>
<td>202 (51.3)</td>
</tr>
<tr>
<td>Girls</td>
<td>192 (48.7)</td>
</tr>
<tr>
<td><strong>Age of the child</strong></td>
<td></td>
</tr>
<tr>
<td>4-5 year</td>
<td>126 (32.0)</td>
</tr>
<tr>
<td>6-7 years</td>
<td>124 (31.5)</td>
</tr>
<tr>
<td>8-10 years</td>
<td>144 (35.6)</td>
</tr>
<tr>
<td><strong>Ethnic background</strong></td>
<td></td>
</tr>
<tr>
<td>Dutch</td>
<td>373 (94.7)</td>
</tr>
<tr>
<td>Non-Dutch</td>
<td>5 (1.4)</td>
</tr>
<tr>
<td><strong>Education Father</strong></td>
<td></td>
</tr>
<tr>
<td>Lower education</td>
<td>79 (20.0)</td>
</tr>
<tr>
<td>Medium education</td>
<td>140 (35.5)</td>
</tr>
<tr>
<td>Higher education</td>
<td>141 (35.8)</td>
</tr>
<tr>
<td><strong>Education Mother</strong></td>
<td></td>
</tr>
<tr>
<td>Lower education</td>
<td>73 (18.5)</td>
</tr>
<tr>
<td>Medium education</td>
<td>179 (45.4)</td>
</tr>
<tr>
<td>Higher education</td>
<td>113 (28.7)</td>
</tr>
</tbody>
</table>

Missing data: ethnic background n = 16, education father n = 34, education mother n = 29
Lower education: no education, primary education or pre-vocational education; Medium education: secondary vocational education; Higher education: university or higher vocational education.
<table>
<thead>
<tr>
<th></th>
<th>Complete N=387</th>
<th>Boys N=199</th>
<th>Girls N=188</th>
<th>4-5 years N=122</th>
<th>6-7 years N=124</th>
<th>8-10 years N=141</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>P90</td>
<td>Mean (SD)</td>
<td>P90</td>
<td>Mean (SD)</td>
<td>P90</td>
</tr>
<tr>
<td>Emotional Symptoms</td>
<td>1.3 (1.9)</td>
<td>4</td>
<td>1.1 (1.6)</td>
<td>4</td>
<td>1.5 (2.1)</td>
<td>4</td>
</tr>
<tr>
<td>Conduct Problems</td>
<td>0.7 (1.3)</td>
<td>2</td>
<td>1.0 (1.5)</td>
<td>3</td>
<td>0.5 (1.1)</td>
<td>2</td>
</tr>
<tr>
<td>Hyperactivity / Inattention Problems</td>
<td>2.7 (2.8)</td>
<td>7</td>
<td>3.5 (3.0)</td>
<td>8</td>
<td>1.8 (2.3)</td>
<td>5</td>
</tr>
<tr>
<td>Peer Problems</td>
<td>1.2 (1.6)</td>
<td>4</td>
<td>1.4 (1.7)</td>
<td>4</td>
<td>1.0 (1.4)</td>
<td>3</td>
</tr>
<tr>
<td>Prosocial Behavior</td>
<td>8.0 (2.3)</td>
<td>10</td>
<td>7.2 (2.4)</td>
<td>10</td>
<td>8.8 (1.8)</td>
<td>10</td>
</tr>
<tr>
<td>Total Difficulties Score</td>
<td>5.9 (5.2)</td>
<td>14</td>
<td>7.1 (5.5)</td>
<td>15</td>
<td>4.7 (4.6)</td>
<td>11</td>
</tr>
<tr>
<td>Impact score</td>
<td>0.4 (0.9)</td>
<td>2</td>
<td>0.6 (1.1)</td>
<td>3</td>
<td>0.2 (0.6)</td>
<td>1</td>
</tr>
</tbody>
</table>

*p90 score = 90th percentile score recommended for clinical use"
Chapter 4

Reliability and Concurrent Validity of the SDQ-T

In Table 3 the internal consistency of the SDQ-T is presented. Especially the hyperactivity / inattention, prosocial and total difficulties scales of the SDQ-T all had Cronbach’s α of 0.80 or above 0.80. The Cronbach’s α were low for conduct (0.64) and peer problems (0.60) and the impact score (0.48).

Table 3 Internal consistency of the SDQ-T

<table>
<thead>
<tr>
<th>SDQ-T subscales</th>
<th>Cronbach’s α</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotional Symptoms</td>
<td>0.73</td>
</tr>
<tr>
<td>Conduct Problems</td>
<td>0.64</td>
</tr>
<tr>
<td>Hyperactivity / Inattention</td>
<td>0.86</td>
</tr>
<tr>
<td>Peer Problems</td>
<td>0.60</td>
</tr>
<tr>
<td>Prosocial Behaviour</td>
<td>0.81</td>
</tr>
<tr>
<td>Total Difficulties</td>
<td>0.80</td>
</tr>
<tr>
<td>Impact Score</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Missing data on the impact score n=12

As demonstrated in Table 4, concurrent validity was shown by a strong (> 0.5) and significant correlation of all subscales of the SDQ-T with the corresponding scale on the TRF except for peer problems. The peer problem scale showed an acceptable significant correlation of 0.46. The prosocial subscale did have a negative correlation with both Internalizing (-0.16) and Externalizing (-0.43) broadband scale of the TRF. The highest correlation (0.73) was found between the SDQ-T TDS and TRF TPS.
### Table 4: Concurrent validity (correlation coefficient) with the TRF of the teacher SDQ

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SDQ-T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotional Symptoms</td>
<td>0.52**</td>
<td>0.44**</td>
<td>0.26**</td>
<td>0.44**</td>
<td>0.36**</td>
<td>0.26**</td>
<td>-0.05</td>
<td>0.003</td>
<td>0.54**</td>
<td>0.28</td>
<td>0.40**</td>
</tr>
<tr>
<td>Conduct Problems</td>
<td>0.10</td>
<td>0.02</td>
<td>-0.01</td>
<td>0.23**</td>
<td>0.17**</td>
<td>0.31**</td>
<td>0.42**</td>
<td>0.59**</td>
<td>0.05</td>
<td>0.57**</td>
<td>0.42**</td>
</tr>
<tr>
<td>Hyperactivity/Inattention</td>
<td>0.12</td>
<td>0.12</td>
<td>0.01</td>
<td>0.33**</td>
<td>0.27**</td>
<td>0.67**</td>
<td>0.33**</td>
<td>0.52**</td>
<td>0.10</td>
<td>0.51**</td>
<td>0.59**</td>
</tr>
<tr>
<td>Peer Problems</td>
<td>0.39**</td>
<td>0.38**</td>
<td>0.10</td>
<td><strong>0.46</strong></td>
<td>0.36**</td>
<td>0.32**</td>
<td>0.15**</td>
<td>0.26**</td>
<td>0.38**</td>
<td>0.24**</td>
<td>0.44**</td>
</tr>
<tr>
<td>Prosocial Behavior</td>
<td>-0.14**</td>
<td>-0.22**</td>
<td>-0.03</td>
<td>-0.24**</td>
<td>-0.16**</td>
<td>-0.33**</td>
<td>-0.32**</td>
<td>-0.40**</td>
<td>-0.16**</td>
<td>-0.43**</td>
<td>-0.43**</td>
</tr>
<tr>
<td>Total Difficulties</td>
<td>0.41**</td>
<td>0.36**</td>
<td>0.14**</td>
<td>0.55**</td>
<td>0.42**</td>
<td>0.65**</td>
<td>0.32**</td>
<td>0.51**</td>
<td>0.38**</td>
<td>0.51**</td>
<td>0.73**</td>
</tr>
</tbody>
</table>

**Significant at 0.01 level (Spearman’s correlation coefficient). Corresponding scales are printed in bold. SDQ-T: Teacher version of the Strengths and Difficulties Questionnaire. TRF: Teacher’s Report Form. The TRF Internalizing Problems broadband scale consists of the Anxious/Depressed,Withdrawn/Depressed and Somatic Complaints syndrome scales. The TRF Externalizing Problems broadband scale consists of the Rule-Breaking Behavior and Aggressive Behavior syndrome scales.**
Using a SDQ-T cut-off score > 14, the SDQ-T had a good sensitivity 90% (95% confidence interval (CI) = 0.86-0.94), specificity 94% (95% CI = 0.92-0.94) and a positive predictive value (PPV) of 0.30 (95% CI = 0.26-0.34). The area under the curve (AUC) using the TRF TPS as criterion was 0.95 (95% CI = 0.89-1.00).

**DISCUSSION**

Our study demonstrated that the SDQ-T is a reliable and valid instrument for identifying emotional and behavioral problems by teachers in a broad age range of school children. The results of this study contribute to the literature about the validity of the SDQ-T in 4-10 year old children. The good psychometric properties and brevity of the SDQ-T make it an easily applicable alternative for the TRF to obtain information about emotional and behavioral problems from teachers on primary school children.

Our findings on the reliability revealed a very good internal consistency of the SDQ-T TDS (α =0.80) and of the subscales hyperactivity / inattention and prosocial problems (α = 0.86 and α = 0.81 respectively). A low reliability of some subscales of the SDQ-T in our study (conduct problems (α =0.64), peer problems (α =0.60) and the impact score (α =0.48), has been reported in two other studies and in studies of the parent version of the SDQ (SDQ-P). Theunissen et al. (2013) concluded in their study of SDQ-P in preschool children (3-4 years) that the low internal consistency of the SDQ subscales does not justify the use of these subscales to decide on a specific need of individual children for further attention regarding these problems. This can be concluded for the subscales of the SDQ-T as well. However an exception could possibly be made for the hyperactivity / inattention problem scale of the SDQ-T; this subscale demonstrated both the highest reliability (0.86) and highest validity (0.67) in our study. These findings are consistent with the study of Posserud et al (2013) that demonstrated high sensitivity of the SDQ for Attention Deficit Hyperactivity Disorder.

Concurrent validity of all subscales, except the peer problems scale, of the SDQ-T with the corresponding scale on the TRF (range 0.54-0.73) was good. Our correlation results were comparable with the results found by the study of Mieloo et al in 5-year old children (correlation range: 0.43-0.76). The relatively low correlation between the peer problems scale with the corresponding TRF social problem scale, was also reported in the study of Mieloo (2012) and Van Leeuwen (2006). An explanation might be that teachers are considered outsiders to the peer group of children and use an adult perspective to interpret children’s social interactions, which decreases the degree of consistency across various scales.

The p90 scores for the teacher report of the SDQ in the 5-year old children in the study of Mieloo et al were remarkably lower than the 4-5 year old p90 scores in our study (p90
How reliable and valid is the teacher version of the SDQ

score SDQ-T TDS 11 versus p90 score SDQ-T TDS 15 respectively). One reason for this difference could be that we included also 4-year-old children in our analysis in this age group. The clinical cut-off point for scoring the SDQ-T TDS in 2-4 years old children in United Kingdom was also 15.

The sensitivity and specificity of the SDQ-T are slightly higher than found for the SDQ-P. In our study the cut-off score of the SDQ-T was higher than the SDQ-P cut-off score measured in Dutch parents. This was also reported in a large Danish study. One explanation could be that teachers may be influenced by some sort of “halo-effect” which means that children exhibiting problem behavior in one area are more likely to be rated as problematic in other areas as well, due to the impact of one class of behavior on the perception of another one.

**Strengths and Limitations**

The current study is, to our knowledge, the first to assess the concurrent validity of the SDQ-T compared to the TRF as a gold standard in a large community sample of 4-10 year old school children. This study did not assess construct validity of the teacher version of the SDQ since multiple other studies already identified evidence for a five-factor model of the SDQ-T. A limitation is the relatively low response rate for informed consent by schools and parents. No data was collected to objectively measure if the schools that refused to participate in our study were in any way different than the schools that chose to participate in our study. A second limitation is that children from immigrant origin and parents with lower vocational education were underrepresented in this study. This may have led to an underestimation of the prevalence of clinical SDQ and TRF scores but it is unlikely that this has significantly influenced the correlation of the SDQ-T with the TRF. A third limitation is the possible clustering of student evaluations by classroom and thus by teacher. However, each teacher assessed only 5 children, largely reducing the potential influence of this. A final limitation is that although the TRF is one of the best instruments available, it cannot be regarded as the ultimate gold standard, because that position is reserved for clinical diagnosis. Because of complexity and high costs, structured clinical interviews such as the Diagnostic Interview Schedule for Children were not used as criterion.

**IMPLICATIONS**

The SDQ-T is a reliable and valid instrument for identifying emotional and behavioral problems in primary school children. The brevity of the SDQ-T makes it an easily applicable questionnaire for obtaining information about emotional and behavioral problems on primary school children from teachers. Obtaining teacher’s SDQ ratings in addition to parental information is valuable for clinicians because this provides data on emotional and
Chapter 4

behavioral problems in children in a second setting, and can be of help in the management of mental health problems in these children.\textsuperscript{13,35} The SDQ-T TDS may highly add to the identification of emotional and behavioral problems in the school setting in these children.
REFERENCES


