

The Strengths and Difficulties Questionnaire: A Research Note

Robert Goodman

Institute of Psychiatry, London, U.K.

A novel behavioural screening questionnaire, the Strengths and Difficulties Questionnaire (SDQ), was administered along with Rutter questionnaires to parents and teachers of 403 children drawn from dental and psychiatric clinics. Scores derived from the SDQ and Rutter questionnaires were highly correlated; parent–teacher correlations for the two sets of measures were comparable or favoured the SDQ. The two sets of measures did not differ in their ability to discriminate between psychiatric and dental clinic attenders. These preliminary findings suggest that the SDQ functions as well as the Rutter questionnaires while offering the following additional advantages: a focus on strengths as well as difficulties; better coverage of inattention, peer relationships, and prosocial behaviour; a shorter format; and a single form suitable for both parents and teachers, perhaps thereby increasing parent–teacher correlations.

Keywords: Questionnaire, child behaviour, psychopathology, strengths.

Abbreviations: CBCL: Child Behavior Checklist; ROC: Receiver Operating Characteristic; SDQ: Strengths and Difficulties Questionnaire.

Introduction

This paper describes a brief behavioural screening questionnaire that provides balanced coverage of children and young people's behaviours, emotions, and relationships. The value of this novel Strengths and Difficulties Questionnaire (SDQ) is evaluated against the benchmark set by the Rutter parent and teacher questionnaires. The SDQ has been designed to meet the needs of researchers, clinicians, and educationalists.

The Rutter questionnaires are long-established and highly respected behavioural screening questionnaires that have proved valid and reliable in many contexts (Elander & Rutter, 1996). Though substantially shorter and therefore quicker to complete than the Child Behavior Checklist (CBCL; Achenbach, 1991a), the Rutter parent questionnaire seem no less useful for many purposes (Berg, Lucas, & McGuire, 1992; Elander & Rutter, 1995; Fombonne, 1989). Developed three decades ago, the Rutter questionnaires have generally worn well, though they do show their age in some ways. Thus all items are about undesirable traits whereas the recent trend, particularly in education, has been to emphasise children's strengths and not just their deficits. In addition, the range of behavioural items covered by the Rutter questionnaires is now somewhat dated. Thus nail-biting and thumb-sucking are included whereas many areas of contemporary interest—including concentration,

impulsivity-reflectiveness, having friends, being victimised, and acting prosocially—are poorly covered. Finally, whereas one version of the Achenbach questionnaire is designed for completion by young people themselves (Achenbach, 1991b), there is no equivalent Rutter questionnaire for self-completion.

A previous research note (Goodman, 1994) described an expanded Rutter parent questionnaire that incorporated all of the original Rutter items as well as many additional items, mostly on children's strengths. The inclusion of these additional items did not appear to attenuate the valuable properties of the original Rutter questionnaire as a behaviour screening instrument, though the extra items presumably did make the questionnaire somewhat more time-consuming to complete. Factor analyses suggested that among children of normal intelligence the expanded questionnaire was tapping five distinct dimensions: conduct problems, emotional symptoms, hyperactivity, peer problems, and prosocial behaviour.

Using these findings as a guide, the SDQ was designed to meet the following specifications: it should fit easily on one side of paper; it should be applicable to children and young people ranging from 4 to 16 years; the same version should be completed by parents and teachers; a similar version should be available for self-report; both strengths and difficulties should be well represented; and there should be equal numbers of items on each of five relevant dimensions, namely conduct problems, emotional symptoms, hyperactivity, peer relationships, and prosocial behaviour. This paper compares informant-completed SDQs with Rutter parent and teacher questionnaires.

Requests for reprints, or sample questionnaires (available in many languages), to: Dr. R. Goodman, Department of Child and Adolescent Psychiatry, Institute of Psychiatry, De Crespigny Park, London SE5 8AF, U.K.

Materials and Methods

Sampling

Questionnaires were obtained on 403 children aged 4–16 years attending one of two London child psychiatric clinics or the children's department of a London dental hospital. The parents of children attending these clinics were recruited into the study until a planned total of roughly 150–250 children had been attained for both dental and psychiatric samples. In the dental clinic and one of the psychiatric clinics (Clinic A), parents who had given informed consent were asked to complete two behavioural screening questionnaires while awaiting their clinic appointment. Participating parents were subsequently asked for permission for their child's teacher to be approached on a similar basis. The other psychiatric clinic (Clinic B) routinely used questionnaires prior to the first assessment, sending them to all parents and, when permission was obtained, to teachers as well. In this clinic, parents were routinely sent both behavioural screening questionnaires and asked if they would be willing for their answers (and the teacher's answers) to be used not only for clinical purposes but also for research. Some of the parents from Clinic B did not complete questionnaires themselves but did give permission for teacher questionnaires to be used for research. The proportion of refusals was not systematically recorded since, as explained later, the statistical analyses did not require the samples to be representative.

Methods

Respondents were administered a Rutter questionnaire and a Strengths and Difficulties Questionnaire (SDQ) in randomised order. Parents were given the Rutter A(2) Questionnaire and teachers the Rutter B(2) Questionnaire; both were scored in the standard way to generate scores for total deviance, conduct problems, emotional symptoms, and hyperactivity (Rutter, 1967; Rutter, Tizard, & Whitmore, 1970; Schachar, Rutter, & Smith, 1981).

The informant-rated version of the SDQ was administered to both parents and teachers. This version of the SDQ is reproduced in full in Appendix A for information only. The SDQ asks about 25 attributes, 10 of which would generally be thought of as strengths, 14 of which would generally be thought of as difficulties, and one of which—"gets on better with adults than with other children"—is neutral. Though no SDQ item is identically worded to any Rutter item, five items are similarly worded. The initial choice of items was guided by the factor loadings and frequency distributions that had previously been obtained on an expanded Rutter parent questionnaire (Goodman, 1994); items were subsequently modified and amalgamated on the basis of a succession of informal trials as well as advice from colleagues. The 25 SDQ items are divided between 5 scales of 5 items each, as shown below.

Hyperactivity Scale. "Restless, overactive, cannot stay still for long"; "Constantly fidgeting or squirming"; "Easily distracted, concentration wanders"; "Thinks things out before acting"; and "Sees tasks through to the end, good attention span".

Emotional Symptoms Scale. "Often complains of headaches, stomach-ache or sickness"; "Many worries, often seems worried"; "Often unhappy, down-hearted or tearful"; "Nervous or clingy in new situations, easily loses confidence"; and "Many fears, easily scared".

Conduct Problems Scale. "Often has temper tantrums or hot tempers"; "Generally obedient, usually does what adults request"; "Often fights with other children or bullies them"; "Often lies or cheats"; and "Steals from home, school or elsewhere".

Peer Problems Scale. "Rather solitary, tends to play alone"; "Has at least one good friend"; "Generally liked by other children"; "Picked on or bullied by other children"; and "Gets on better with adults than with other children".

Prosocial Scale. "Considerate of other people's feelings"; "Shares readily with other children (treats, toys, pencils, etc.)"; "Helpful if someone is hurt, upset or feeling ill"; "Kind to younger children"; and "Often volunteers to help others (parents, teachers, other children)".

Each item can be marked "not true", "somewhat true" or "certainly true". For all of the items except the five printed above in italics, the item is scored 0 for "not true", 1 for "somewhat true", and 2 for "certainly true". For the five items printed above in italics, the item is scored 2 for "not true", 1 for "somewhat true", and 0 for "certainly true". The score for each of the five scales is generated by summing the scores for the five items that make up that scale, thereby generating a scale score ranging from 0 to 10. The scores for hyperactivity, emotional symptoms, conduct problems, and peer problems can be summed to generate a total difficulties score ranging from 0 to 40; the prosocial score is not incorporated in the reverse direction into the total difficulties score since the absence of prosocial behaviours is conceptually different from the presence of psychological difficulties.

The Rutter A(2) and the SDQ were both completed by the parents of 346 children: 158 dental clinic attenders and 188 psychiatric clinic attenders. The Rutter B(2) and the SDQ were both completed by the teachers of 185 children: 39 dental clinic attenders and 146 psychiatric clinic attenders. Most of the teacher reports were on psychiatric clinic attenders because the parents of children attending Child Psychiatric Clinic B generally agreed to the clinic sending questionnaires to teachers for clinical as well as research purposes; parents of children attending the dental clinic or Child Psychiatric Clinic A were less likely to give permission for teacher questionnaires to be obtained solely for research purposes.

Statistical Analysis

As in previous studies comparing the validity of different screening questionnaires (e.g. Berg et al., 1992), analyses of Receiver Operating Characteristic (ROC) curves were used to establish how well each questionnaire was able to distinguish between high- and low-risk samples, determining the area under the curve for each questionnaire (Hanley & McNeil, 1982). For this purpose, the only underlying assumption is that children recruited from the two psychiatric clinics were substantially more likely to have psychiatric disorders than children recruited from the dental clinic. There is no assumption that all subjects recruited from the psychiatric clinics had psychiatric disorders, nor that all subjects recruited from the dental clinic were free from psychiatric disorder. Equally, there is no assumption that the psychiatric sample was representative of all children between 4 and 16 who attend psychiatric clinics, nor that the dental sample was representative of all children attending dental clinics, let alone of all children aged between 4 and 16. Since the ROC curves for the SDQ and Rutter questionnaires were derived from the same set of patients, the statistical comparison of the areas under these ROC curves allowed for the paired nature of the data (Hanley & McNeil, 1982). Comparison of the parent-teacher correlations of the SDQ and Rutter questionnaires also allowed for the paired nature of the data, using structural equation modelling (EQS, BMDP Statistical Software) and examining whether constraining the two correlations to be the same resulted in a significantly poorer fit. Though appropriate for tests of comparative validity and cross-situation correlation, the case-control sampling used in this study does not generate sensitivity or specificity estimates that

could securely be generalised to representative epidemiological or clinical samples; such estimates will subsequently be derived from other studies in progress.

Reported correlations are Pearson product-moment correlations, but the pattern of findings was not changed when Spearman correlations were used instead. Intraclass correlations—which are often appropriate for reliability estimates—were not used to measure parent–teacher agreement even though this agreement could be construed as an index of inter-rater reliability. Parents and teachers make ratings based on different sources of information, whereas measures of inter-rater reliability are more appropriately derived from independent ratings based on the same source of information. Furthermore, employing intraclass correlations would have involved mixing parent- and teacher-derived scores, and this would have been inappropriate since mean scores differed systematically between parent and teacher ratings—a difference allowed for when interpreting these scores (Rutter, 1967; Rutter et al., 1970; and see Appendix B).

Results

Age and Gender

The mean age (*SD*) of the dental sample was 10.8 years (3.1) while that of the psychiatric sample was 9.8 years (3.3), a significant difference [$t(401) = 3.00, p < .01$]. As expected, the proportion of males was higher in the psychiatric sample (63%, 153/244) than in the dental sample (53%, 85/159) [continuity-adjusted $\chi^2(1) = 3.03, p < .05$, 1-tailed]. The results reported here are for the sample as a whole, though closely similar results were obtained when ROC and correlational analyses were repeated separately for boys and girls, and separately for children aged 4–10 and 11–16.

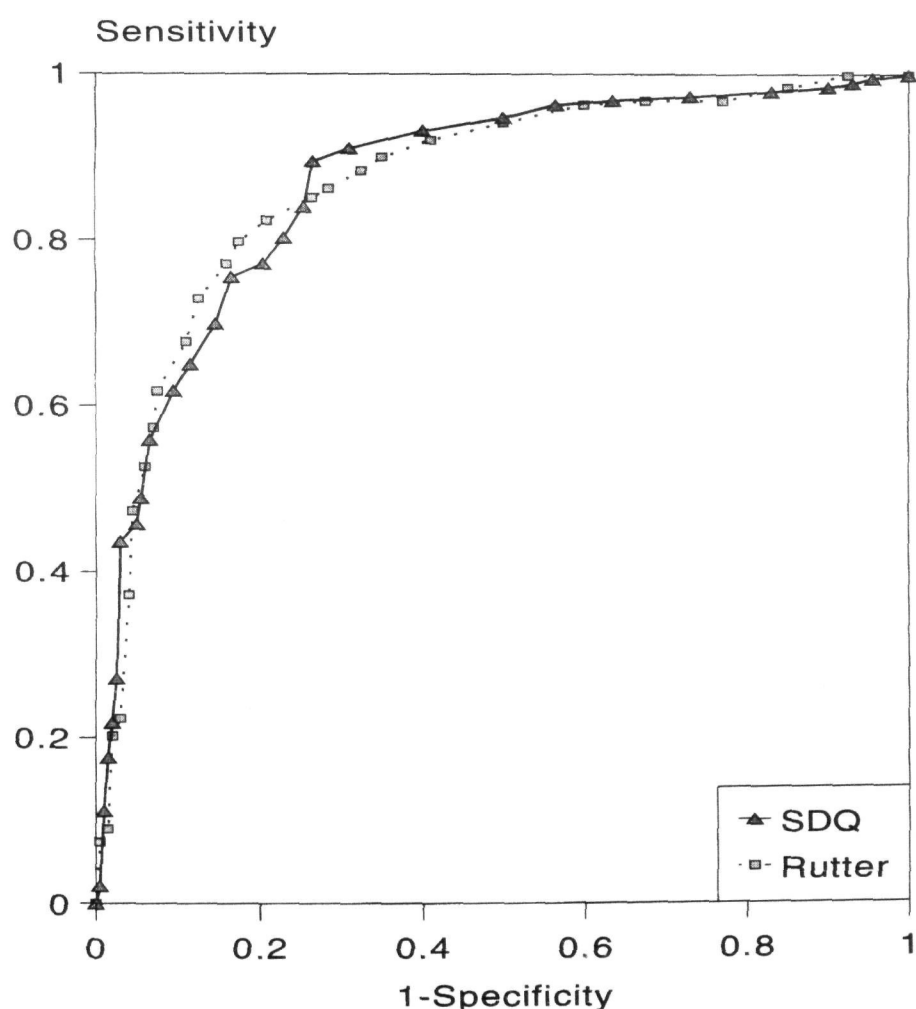


Figure 1. ROC curves for parent-rated questionnaires.

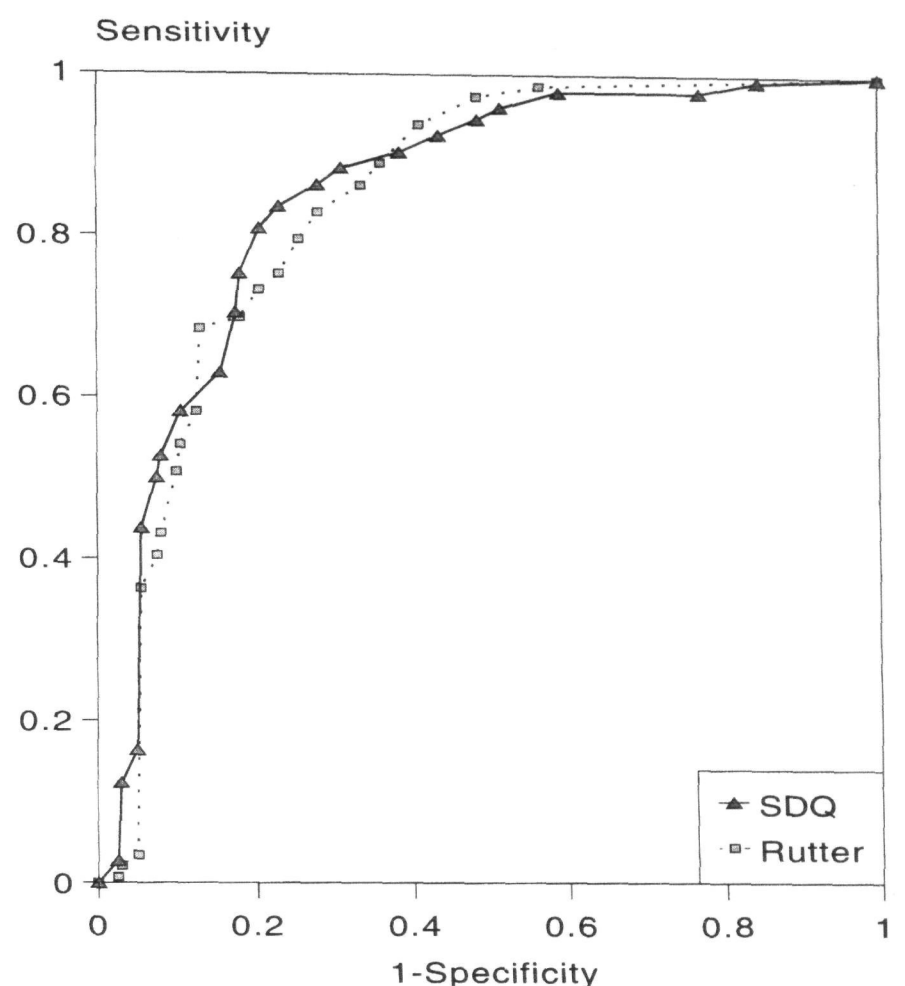


Figure 2. ROC curves for teacher-rated questionnaires.

Discriminating between Psychiatric and Nonpsychiatric Samples

The ability of the two questionnaires to distinguish between dental and psychiatric cases is reflected in the Receiver Operating Characteristic (ROC) curves shown in Figs. 1 and 2 for parent and teacher reports respectively. The ROC curves for the Rutter questionnaires are based on total deviance scores, whereas the SDQ curves are based on total difficulties scores. The comparable ability of the two measures to discriminate between the two samples is evident from the extent to which the two curves almost superimpose on one another. Quantitatively, this comparability can be judged from the area under each of the curves, which is a measure of how well that measure discriminates between the two samples; the area under the curve would be 1.0 for a measure that discriminated perfectly, and .5 for a measure that discriminated with no better than chance accuracy. For parent reports, the area under the curve (95% confidence interval) was .87 (.83–.91) for the SDQ as compared with .87 (.83–.91) for the Rutter A(2) parent questionnaire—a nonsignificant difference ($z = .13, p = .9$). For teacher reports, the corresponding values were .85 (.78–.93) for the SDQ as compared with .84 (.76–.93) for the Rutter B(2) teacher questionnaire—a nonsignificant difference ($z = .41, p = .7$).

SDQ–Rutter Correlations

Table 1 shows the correlations between SDQ and Rutter scores. The correlations were only marginally lower when the analyses were repeated for the dental and psychiatric samples separately. No cross-measure corre-

Table 1
Inter-measure Correlation for Each Type of Rater

	SDQ–Rutter correlation	
	Parent report N = 346	Teacher report N = 185
Total Deviance/Difficulties score	.88	.92
Conduct Problems score	.88	.91
Emotional Symptoms score	.78	.87
Hyperactivity score	.82	.90

Table 2
Inter-rater Correlations for Each Type of Measure

	Parent–Teacher correlation (N = 128)	
	SDQ	Rutter
Total Deviance/Difficulties score	.62*	.52
Conduct Problems score	.65	.57
Emotional Symptoms score	.41	.47
Hyperactivity score	.54	.55
Peer Problems score	.59	—
Prosocial Behaviour score	.37	—

* Correlation significantly higher than the comparable Rutter correlation ($p < .02$); all other comparisons non-significant.

lations can be presented on two SDQ scores—the peer problems score and the prosocial behaviour score—since they have no Rutter counterpart.

Parent–Teacher Correlations

Table 2 presents the correlation coefficients between teacher- and parent-derived scores when both are using the SDQ or when both are using Rutter questionnaires. For comparable scores, the cross-situation correlations of the SDQ and Rutter measures were generally similar, apart from the higher SDQ correlation for total score [χ^2 (1) = 5.90, $p < .02$]. Though the correlations were lower when the analyses were repeated for the dental and psychiatric samples separately, these correlations were generally comparable for the SDQ and Rutter measures, apart from a higher SDQ correlation for total score in the psychiatric sample [χ^2 (1) = 4.05, $p < .05$].

Discussion

Given the well-established validity and reliability of the Rutter questionnaires (Elander & Rutter, 1996), the high correlation between the total scores generated by the SDQ and Rutter questionnaires is evidence for the concurrent validity of the SDQ. Parent–teacher correlations were either equivalent for the two measures or slightly favoured the SDQ, perhaps because the SDQ

used identical items for parents and teachers whereas the Rutter questionnaires were somewhat different for parents and teachers. The ROC analyses showed that the two measures had equivalent predictive validity, as judged by their ability to distinguish between psychiatric and nonpsychiatric samples. Of course, discriminating between psychiatric and dental clinic attenders is a relatively easy task, but the high correlation between SDQ and Rutter scores within each clinic group suggests that the two measures are also likely to be comparably discriminating in more demanding screening tasks, such as detecting nonreferred cases of child mental health problems in the community; further empirical studies would be needed to confirm this. Since previous studies have shown that CBCL and Rutter parent questionnaire scores are highly correlated (Berg et al., 1992; Fombonne, 1989), and that these two sets of questionnaires are of comparable predictive validity (Berg et al., 1992), it is likely that the SDQ and CBCL will also be highly correlated and have comparable validity; direct comparisons are currently under way.

The SDQ and Rutter questionnaires can each be used to generate separate scores for conduct problems, emotional symptoms, and hyperactivity. For each of these three scores, there was a high correlation between the SDQ score and the Rutter score; and parent–teacher correlations were comparable for the two sets of measures. Despite its brevity, the SDQ also generated two scores that have no Rutter counterparts: a peer problems score and a prosocial behaviour score.

The performance of the SDQ could potentially have been undermined by three of its design features: inclusion of strengths as well as difficulties; use of an identical questionnaire for both parents and teachers; and a compact presentation on just one side of paper. The equivalence of the SDQ and Rutter scores suggests that these three features have not had an adverse effect. This should encourage researchers and clinicians who are contemplating incorporating similar features into other questionnaires.

Rutter questionnaires are routinely used to categorise children as likely psychiatric “cases” or “non-cases” according to whether their total deviance score is equal to or greater than a standard cut-off: 13 on the Rutter parent questionnaire and 9 on the Rutter teacher questionnaire (Rutter, 1967; Rutter et al., 1970). Using a single cut-off for all studies has both advantages and disadvantages. The advantages are simplicity and equivalence across studies. The main disadvantage is that “caseness” does not have a comparable meaning in

different studies simply because those studies have employed the same cut-off. Comparability is particularly likely to be lost when high- and low-risk samples are contrasted. A worked example may be helpful. Study X involves 100 children from a high-risk population with a true rate of psychiatric disorder of 50%; if the screening questionnaire has a sensitivity of .8 and a specificity of .8 when using the standard cut-off, the questionnaire will identify 40 true positives and 10 false positives. Study Y involves 100 children from a low-risk population with a true rate of disorder of 10%; even with the same sensitivity and specificity, the questionnaire will identify 8 true positives and 18 false positives. Despite using the same questionnaire and the same cut-off, a comparison of "cases" from studies X and Y will primarily be a comparison of true positives from study X with false positives from study Y.

Given these problems, the best strategy for researchers may be to choose cut-offs according to the likely disorder rate in the sample being studied, and according to the relative importance for that study of false positives and false negatives. It may also be appropriate to adjust cut-offs for age and gender. Ongoing clinical and epidemiological studies using the SDQ should provide the basis for cut-offs adjusted for these sample characteristics. In addition, planned trials should establish if the predictive validity of the SDQ can further be improved by an algorithm that combines SDQ scores with scores from an additional and even briefer screening instrument that elicits the respondent's view on whether the child has significant emotional or behavioural difficulties, and on the extent to which these difficulties result in social impairment or distress for the child, or burden for others. Until these various studies are completed, SDQ users can use the provisional cut-off scores shown in Appendix B, which are derived partly from the samples used for this study and partly from other ongoing epidemiological surveys using the SDQ. The "borderline" cut-offs can be used for studies of high-risk samples where false positives are not a major concern; the "abnormal" cut-offs can be used for studies of low-risk samples where it is more important to reduce the rate of false positives.

Conclusion

These initial findings suggest that the SDQ may function as well as the Rutter questionnaires (and, by inference, the Achenbach questionnaires) while offering

the following additional advantages: a compact format; a focus on strengths as well as difficulties; better coverage of inattention, peer relationships, and prosocial behaviour; and a single form suitable for both parents and teachers, perhaps thereby increasing parent-teacher correlations.

Acknowledgements—I am very grateful for the willing co-operation of parents and teachers, and for the invaluable assistance of Dr. Hilary Richards and the staff of the three London clinics that participated in the study: the Department of Paediatric Dentistry of King's Dental Institute, Camberwell Child Guidance Centre, and the Department of Child and Adolescent Psychiatry, Hounslow.

References

- Achenbach, T. M. (1991a). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington, VT: University of Vermont Department of Psychiatry.
- Achenbach, T. M. (1991b). *Manual for the Youth Self-Report and 1991 Profile*. Burlington, VT: University of Vermont Department of Psychiatry.
- Berg, I., Lucas, C., & McGuire, R. (1992). Measurement of behaviour difficulties in children using standard scales administered to mothers by computer: Reliability and validity. *European Child and Adolescent Psychiatry*, 1, 14-23.
- Elander, J., & Rutter, M. (1996). Use and development of the Rutter Parents' and Teachers' Scales. *International Journal of Methods in Psychiatric Research*, 6, 63-78.
- Fombonne, E. (1989). The Child Behavior Checklist and the Rutter Parental Questionnaire: A comparison between two screening instruments. *Psychological Medicine*, 19, 777-785.
- Goodman, R. (1994). A modified version of the Rutter parent questionnaire including items on children's strengths: A research note. *Journal of Child Psychology and Psychiatry*, 35, 1483-1494.
- Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143, 29-36.
- Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, 148, 839-843.
- Rutter, M. (1967). A children's behaviour questionnaire for completion by teachers: Preliminary findings. *Journal of Child Psychology and Psychiatry*, 8, 1-11.
- Rutter, M., Tizard, J., & Whitmore, K. (1970). *Education, health and behaviour*. London: Longman.
- Schachar, R., Rutter, M., & Smith, A. (1981). The characteristics of situationally and pervasively hyperactive children: Implications for syndrome definition. *Journal of Child Psychology and Psychiatry*, 22, 375-392.

Accepted manuscript received 26 September 1996

Appendix A: Strengths and Difficulties Questionnaire

For each item, please mark the box for Not True, Somewhat True or Certainly True. It would help us if you answered all items as best you can even if you are not absolutely certain or the item seems daft! Please give your answers on the basis of the child's behaviour over the last six months or this school year.

Child's Name

Male / Female

Date of Birth.....

	Not True	Somewhat True	Certainly True
Considerate of other people's feelings	☐	☐	☐
Restless, overactive, cannot stay still for long	☐	☐	☐
Often complains of headaches, stomach-aches or sickness	☐	☐	☐
Shares readily with other children (treats, toys, pencils etc.)	☐	☐	☐
Often has temper tantrums or hot tempers	☐	☐	☐
Rather solitary, tends to play alone	☐	☐	☐
Generally obedient, usually does what adults request	☐	☐	☐
Many worries, often seems worried	☐	☐	☐
Helpful if someone is hurt, upset or feeling ill	☐	☐	☐
Constantly fidgeting or squirming	☐	☐	☐
Has at least one good friend	☐	☐	☐
Often fights with other children or bullies them	☐	☐	☐
Often unhappy, down-hearted or tearful	☐	☐	☐
Generally liked by other children	☐	☐	☐
Easily distracted, concentration wanders	☐	☐	☐
Nervous or clingy in new situations, easily loses confidence	☐	☐	☐
Kind to younger children	☐	☐	☐
Often lies or cheats	☐	☐	☐
Picked on or bullied by other children	☐	☐	☐
Often volunteers to help others (parents, teachers, other children)	☐	☐	☐
Thinks things out before acting	☐	☐	☐
Steals from home, school or elsewhere	☐	☐	☐
Gets on better with adults than with other children	☐	☐	☐
Many fears, easily scared	☐	☐	☐
Sees tasks through to the end, good attention span	☐	☐	☐

Signature

Date

Parent/Teacher/Other (please specify:)

Thank you very much for your help

© Robert Goodman, 1977

Appendix B: Provisional Banding of SDQ Scores

These bands, which are not adjusted for age or gender, have been chosen so that roughly 80 % of children in the community are normal, 10 % are borderline, and 10 % are abnormal.

	Normal	Borderline	Abnormal
Parent completed			
Total Difficulties Score	0–13	14–16	17–40
Emotional Symptoms Score	0–3	4	5–10
Conduct Problems Score	0–2	3	4–10
Hyperactivity Score	0–5	6	7–10
Peer Problems Score	0–2	3	4–10
Prosocial Behaviour Score	6–10	5	0– 4
Teacher completed			
Total Difficulties Score	0–11	12–15	16–40
Emotional Symptoms Score	0–4	5	6–10
Conduct Problems Score	0–2	3	4–10
Hyperactivity Score	0–5	6	7–10
Peer Problems Score	0–3	4	5–10
Prosocial Behaviour Score	6–10	5	0– 4

This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.