

A multiple criterion screen for identifying secondary school children with psychiatric disorder

Characteristics and efficiency of screen

A. MACMILLAN,¹ I. KOLVIN,² R. F. GARSIDE,
A. R. NICOL AND I. M. LEITCH

From the Human Development Unit, University of Newcastle-upon-Tyne

SYNOPSIS A multiple criterion screen was developed as part of a study which was designed to assess different types of intervention for school children with psychiatric disorder. The foremost requirement was that false positives be kept to a minimum, with selected children being unequivocally disturbed. Children were identified for treatment by the following measures: (a) *teacher rating* on the Rutter B2 Scale, giving a total score and subscores for neurotic and antisocial behaviour; (b) *peer ratings* on a sociometric measure, yielding isolation and rejection scores; and (c) *self ratings* on the Junior Eysenck Personality Inventory, specifically the neuroticism dimension. A weighting system was developed so that children could be selected on the basis of extreme scores on either teacher or self-rating measures alone, or by a combination of less extreme scores on more than one measure. The results are presented and discussed.

INTRODUCTION AND AIMS

The design of a population screen should be determined by the use to which it is to be put. Most screens in child psychiatric research have been aimed at establishing prevalence rates and elucidating the relationship of psychiatric disorder to population and environmental characteristics (Mitchell & Shepherd, 1966; Rutter *et al.* 1970). The main aim of the present study was to give help to children with special needs in ordinary schools and the aim of the screen was to identify these children. This difference in aim has a number of implications.

(1) Since the special help was to be given at school, we were primarily interested in disorders which were manifested in the school situation. It was less important that we identified all disorders, regardless of the situation in which they were manifest, than in studies whose prime aim was to establish prevalence rates of disorders.

(2) It was of the greatest importance that we explored in full the various facets of disturbed behaviour that may manifest themselves at school, since this would have implications for assessment and treatment.

(3) It was important that screening be as rapid and economical as possible, so that energy was not deflected from the main purpose of the study, which was to obtain special help for children.

(4) It was important that a high proportion of the cases identified were really suffering from a handicap (i.e. were 'true positives'). The screen was followed by an 'in-depth' investigation of each case, so that there was no danger that an incorrectly identified case would receive treatment. At the same time, it would be a waste of precious resources if a large number of cases were investigated and found to have been wrongly identified (i.e. were 'false positives'). Equally, we wanted to be sure that not too many children in need were missed (i.e. we aimed to keep 'false negatives' to a minimum).

These considerations pointed to the value of a multiple criterion screen (Bower, 1969). This is based entirely in the school and draws on facets of behaviour as seen by the teacher, peers and self.

¹ Address for reprints: Mr A. Macmillan, Department of Psychological Services and Research, Crichton Royal Hospital, Dumfries.

² Address for correspondence: Professor I. Kolvin, Human Development Unit, 1 Tankerville Terrace, Jesmond, Newcastle-upon-Tyne NE2 3AH.

There are difficulties in using teachers' ratings alone, this being the most commonly used screening technique. For example, correctives may be required for the fact that teachers tend to perceive boys' and girls' behaviour differently (Brophy & Good, 1970; Wilson, 1974); also, they may overlook quiet, passive, but potentially disturbed children (Garner & Bing, 1973); and there may be more general differences in the perception of problem behaviour - 'one man's disruption may approximate another's peak teaching experience', as Davies (1976) picturesquely expresses it. In addition, teachers may not be sufficiently sensitized to the interpersonal difficulties experienced by some children, and sociometric information is of considerable value here. Self-ratings by the child serve to complement teacher- and peer-derived information and may reveal the personal unhappiness and discomfort that both other sources bypass. Finally, it should be pointed out that, despite high reliability for a questionnaire, a single rating by a single teacher on such a questionnaire can nevertheless easily over- or under-estimate deviance in an individual child. Completion of successive questionnaires may increase efficiency (Booth & Taylor, 1973).

In this paper we report the development of the multiple criterion screen and the relations between the various measures used. We also report the relations of the measures to the clinical levels of disturbance as derived from more 'in-depth' assessment. This, in turn, allows us to estimate our success in identifying a high proportion of true positives.

We also report the rates of psychiatric disturbance in those children identified as deviant by the screen and also those in a random control group. This allowed us to estimate the population prevalence of psychiatric disturbance.

METHOD

(a) The population

Children in the first year of 6 junior high or comprehensive schools in the north-east of England were screened.¹ This exercise involved 1723 children with a mean age of 11.8 years and

¹ While screening was conducted in 2 consecutive years, the present paper is concerned only with data from the first year.

range 10.2-12.9 years. 97.5% of the children were aged between 11 years 2 months and 12 years 2 months. The number of boys and girls were roughly equal (52.6 and 47.4% respectively). The schools involved were broadly representative of the schools in this area: the social class level of the population in the area is slightly below the national average (Neligan *et al.* 1975).

(b) Screening measures

(i) Teachers' ratings

The scale employed here was the Rutter Teacher Scale B2 (Rutter, 1967). This scale can be quickly completed by teachers with a good knowledge of children's day-to-day behaviour in school. The B version of this scale, differing only in the wording of a few items from the B2, has been demonstrated to have inter-rater reliability of the order of 0.72 and retest reliability of 0.89. As well as a total score, it also yields two subscale scores, for neurotic and antisocial behaviour. The number of items contributing to these 2 subscales was increased in the present study, following an inspection of Rutter's (1967) data bearing on discrimination of the 2 types of disorder. This increase improved the reliability of the subscales. Items were added to the respective subscales on the basis that (i) they differentiated the psychiatric group from the controls; and (ii) they differentiated the diagnostic groups. The revised subscales therefore consist of the following neurotic items: G, H, J, K, N, Q, R, V and W; and conduct items: A, B, D, E, O, P, S, T and Z.

(ii) Sociometry

The development of the sociometric instrument is discussed in detail elsewhere (Macmillan *et al.* 1978). Each child was asked to choose which 3 of his classmates he would like to sit beside and which 3 he would like to play with at play-time and, conversely, those he would *not* like to sit beside or play with. Two scores of interest are derived from this procedure: the first is *isolation*, which is a low score on positive choices; and the second is *rejection*, which is a high score on negative choices. We found that test-retest reliability with a one-month gap was 0.72 for isolation and 0.87 for rejection ($N = 61$).

(iii) *Self-rating*

The measure employed here was the Junior Eysenck Personality Inventory (JEPI) (Eysenck, 1965). This is a 60-item questionnaire, yielding an extroversion-introversion score, a lie score and a neuroticism score. We were particularly concerned with the last of these dimensions. The high scorer on neuroticism is likely to be 'moody, touchy, anxious, restless and rigid' (Eysenck, 1965). These are the characteristics associated with instability, while the stable person (low scorer on neuroticism) is likely to be 'calm, carefree, easy-going and reliable'. Split-half reliabilities for neuroticism with the 11-12 year old group with which we are concerned are in the region of 0.84, while test-retest reliability is reported as lying between 0.53 and 0.87.

(c) *Establishing cut-offs and assigning weightings*

It was decided to take extreme scores on each of the screen measures as indicators of maladjustment. The actual scores used as cut-offs were decided in most cases by examining published data on the characteristics of the instrument. For the subscales of the Rutter questionnaire and sociometry, however, cut-off scores were decided on the basis of a pilot study of 200 cases. With a multiple criterion screen, one can weight each extreme score equally or, alternatively, assign additional weightings to very high scores on particular measures. With the former system, it is theoretically possible for a child to obtain a very deviant score on one specific measure only, but be excluded because his *summed* weighted score is not sufficiently high. To avoid this, we adopted weighting formulae which allowed children with markedly deviant teacher- or self-ratings to be selected on that basis alone.

With the Rutter B2 scale, a cut-off of 9 has been regarded as providing the best discrimination between children attending child guidance clinics and a normal sample (Rutter *et al.* 1970). We retained this cut-off, assigning it a weighting of 1 point towards the deviance classification. In addition, we gave the more extreme score of 15 a weighting of 2. This was arrived at by adding 1 standard deviation to Rutter's original cut-off. A cut-off of 9 identified some 12% of the sample, while a cut-off of 15 identified 2%.

Cut-offs were assigned on the neurotic and antisocial subscales which gave rise to a yield

closest to that of the total score. A score of 4 on the neurotic and antisocial subscales identified 11 and 14% respectively, and hence was assigned a weighting of 1. We should emphasize that our use of the subscale as a weighting score is different from Rutter *et al.*'s use of the subscale for diagnostic purposes. The three Rutter weighted scores were added together to contribute to the total screen score.

With the sociometric criteria, our decisions about cut-offs were guided by our findings with the Rutter scale. Cut-offs were adjusted so that percentages similar to that identified by the Rutter Total cut-off were selected. For isolation, a cut-off of one positive choice or less was selected, this selecting 14% of the pilot children. For rejection, 12 or more negative choices was taken as the cut-off and this also selected 14%. Scores on or beyond these cut-offs were each weighted 1 point.

Cut-offs of $1\frac{1}{2}$ and 2 standard deviations above the mean for neuroticism were taken for the JEPI, with the scores being rounded up to be equivalent for both sexes. These scores were 20 and 23 respectively; 17% of the pilot sample scored $1\frac{1}{2}$ standard deviations above the mean. Children with scores of 20-22 were allotted 2 points, and those with more extreme scores of 23 or 24, 3 points.

The children's scores on each of the screen measures were summed. Those obtaining a total of 3 or more points were regarded as screen positives; those scoring below this total were viewed as screen negatives. From the summary of cut-offs and weighting scores in Table 1, it can be seen that children could be selected as screen positives on the basis of extreme scores on either the Rutter B2 scale or JEPI Neuroticism alone, or else by varying combinations of scores from the 3 instruments. The maximum obtainable weighted score was 9. The weighting score could therefore yield information not only concerning the presence or absence of disturbance but also concerning its severity. Of the 3 screen measures, only the Rutter B2 scale was developed to discriminate between children who show disorder and those who do not. Even so, *all* measures at the screen level can indicate only instrumental deviance; whether this is synonymous with psychiatric disorder is a question for a more detailed level of investigation. One can, however, argue with some measure of confidence

Table 1. *Cut-offs and weighting scores on screen criteria*

Measure	Cut-off	Weighting score
Rutter B2 Scale		
Total score	9-14	1
	15 or above	2
Neurotic subscale (Newcastle modification)	4 or above	1
Antisocial subscale (Newcastle modification)	4 or above	1
Sociometry		
Isolation	0 or 1	1
Rejection	12 or more positive choice negative choices	1
JEPI Neuroticism		
	20-22	2
	23-24	3

that extreme scores on the personality dimension of neuroticism are likely to be associated with some degree of malfunctioning or problems of adjustment, as are scores on sociometric criteria which indicate extreme lack of popularity or rejection.

(d) Selection of controls

Normal controls were selected at random from the residual pool of children who scored less than 3 on the weighting system, with the proviso that they be drawn from all 6 schools, and that they reflect the sex ratio found in the group of screen positives.

(e) Assessment of psychiatric disorder

In order to check the efficiency of the screen criteria in identifying genuinely disturbed children, all the information available on selected children and controls - screen data, more detailed school-based measures and data from parent interview (Kolvin *et al.* 1977) - were examined by 2 psychiatrists (I.K. and A.R.N.), and an assessment of diagnosis, severity and duration of disorder was made. The procedure has been found to have satisfactory reliability, and findings in relation to this exercise will be described later (Wrate *et al.* 1980). With drop-outs due to factors such as parental unwillingness to co-

operate, difficulty in contacting families in order to obtain an interview and children transferring to schools outside the area, the final numbers involved were 180 screen positives and 64 screen negatives. The topic of screen validity is discussed further in a companion paper (Nicol *et al.* 1980).

RESULTS

(a) Numbers selected

The following results are derived from analyses involving 4 groups. These are:

(1) 217 children (12.6% of the population) who were identified as positive on the screen on the basis of scoring 3 points or more on the weighting system described above.

(2) 553 children who scored less than 3 points on the screen. This represented a random 1 in 3 sample of the screen negative children. The sampling was carried out to facilitate data handling.

(3) 180 of the screen positive cases where full psychiatric and social data were collected.

(4) 64 screen negative cases where full psychiatric and social data were collected for control purposes.

The Rutter Total score, with the cut-off employed, selected most cases (147) in the group of screen positives, followed by Rutter Antisocial (128) and Rejection (111). Isolation selected the fewest (64) in this group.

A crude indication of the relative contribution of the criteria in discriminating between screen positives and screen negatives is available from Fig. 1 by examining the ratio of cases falling in the 2 groups. Thus, for Rutter Total the ratio of screen positive to screen negative is 2.6:1; Rutter Neurotic 1.4:1; Rutter Antisocial 1.5:1; JEPI Neuroticism 0.7:1; Isolation 1.1:1; and Rejection 1:1. At one extreme, therefore, much the greater majority of cases identified by Rutter Total is placed in the group of screen positives, while for cases identified by JEPI Neuroticism, rather more fall in the group of screen negatives.

(b) Specificity and overlap in selection

The degree of specificity and overlap of criteria in selection of cases is detailed in Table 2. In

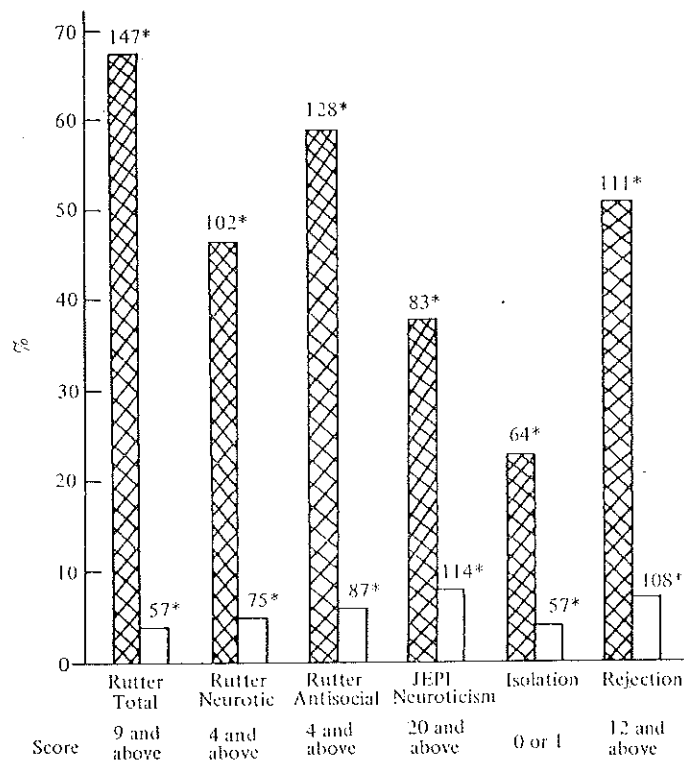


FIG. 1. Percentage of children scoring above the cut-off on each screen criterion out of a total of 217 screen positives and 1506 screen negatives. (Figures for the screen negatives are projected from analysis of a random sample of 553 cases.) ▨, Screen positives; □, screen negatives; *, frequencies.

Table 2. Proportion (%) of cases identified by each main criterion selected by that criterion alone

	Rutter Total	JEPI Neuroticism	Isolation	Rejection
Specific	39	56	14	20
Overlap	61	44	86	80
Number of cases identified	166	125	86	150

In addition to identifying the most children, the Rutter scale has a high degree of overlap with the other criteria. Some 39% of cases identified by this criterion are identified by it alone. Both Isolation and Rejection have high overlap with other criteria and, in the case of Isolation, only 14% are selected by this criterion alone. The JEPI Neuroticism criterion, while identifying a large number of children, had an overlap on only 44% of cases, indicating that, to a much greater extent than other criteria, it is selecting cases not accessible to the other measures.

A more detailed representation of overlap of criteria is given in Table 3. With Rutter Total, most overlap is observed with the Rutter Neurotic and Antisocial scores, not surprisingly, as they contribute to the Total score. The greater overlap with Rutter Antisocial suggests that, in the population we have studied, the total Rutter score reflects antisocial behaviour to a greater extent than neurotic behaviour. The pattern of overlap for all 3 Rutter criteria with the other criteria is similar; it consists in substantial overlap with Rejection, moderate overlap with Isolation, and minimal overlap with JEPI Neuroticism. JEPI Neuroticism overlap is minimal with all criteria other than Rejection, where it reaches moderate (30%) proportions. Turning to sociometric criteria, Isolation and Rejection show a similar pattern of overlap with other criteria, with generally greater overlap with Rutter criteria than with JEPI Neuroticism. A third of cases designated as rejected prove to be isolated. Of the isolated cases, well over half

Table 3. *Overlap (%) of selection criteria in relation to each selection criterion*

Criterion	Number of cases selected by a particular criterion	Rutter Total I	Rutter Neurotic II	Rutter Antisocial III	JEPI Neuroticism IV	Sociometry Isolation V	Sociometry Rejection VI
I	166	100	52	81	11	27	47
II	127	68	100	47	18	28	42
III	159	85	37	100	13	21	45
IV	125	14	18	15	100	18	30
V	86	52	43	40	26	100	59
VI	150	53	36	48	25	33	100

Percentages rounded off to whole numbers.

Table 4. *Correlation of screen criteria*

Screen criterion	Rutter Neurotic	Rutter Antisocial	JEPI Neuroticism	Isolation	Rejection	Total weighted score
Rutter Total	0.46***	0.72***	-0.03	0.21***	0.28***	0.72***
Rutter Neurotic	—	0.21***	0.02	0.19***	0.23***	0.53***
Rutter Antisocial	—	—	-0.03	0.15***	0.26***	0.62***
JEPI Neuroticism	—	—	—	0.05	0.07	0.52***
Isolation	—	—	—	—	0.34***	0.44***
Rejection	—	—	—	—	—	0.55***

$N = 625$ (one-third screen positives + one-third screen negatives). This is based on a representative sample of our population with 625 cases constituting one-third of both the screen positive and screen negative cases.

*** $P < 0.001$.

(59%) prove to be rejected, not surprisingly since the Rejection cut-off selected almost twice as many cases as the Isolation cut-off.

Another way of viewing the relationship between the screen criteria is in terms of their correlations (Table 4). All correlations within and between criteria deriving from teachers and peers are highly significant statistically, although in some cases small (e.g. those involving Rutter Neurotic and Isolation). The most striking aspect of the matrix is that no correlation involving JEPI Neuroticism reaches significance; they are all virtually zero. The most substantial correlations are observed between Rutter Total and the two Rutter subscales, and between the two sociometric criteria. Correlations of moderate extent are observed between Rejection and the 3 Rutter criteria, while the respective correlations for Isolation are among the lowest which reach significance. In terms of correlation with total weighted score, Rutter Total shows the highest correlation, followed by Rutter Antisocial. Rejection, Rutter Neurotic and JEPI Neuroticism each correlate to an almost identical

extent, while Isolation has a lower correlation with all other criteria except JEPI. There is the expected pattern of lower correlations between independent criteria and between each criterion and summed weighted score.

(c) Psychiatric assessment

The results of the psychiatric assessment in which the group of screen positive ($N = 180$) and screen negative ($N = 64$) cases were assessed, without knowledge of screen category, are shown in Table 5. In the total group of screen positives, over half the cases fall in the 'moderate' category of disturbance, the remainder falling into almost equal groups of 'marked' and 'dubious' disturbance. One false positive was recorded. In the total control group, no cases with marked disturbance were found, and the 22 cases where some disturbance was evident divide equally between the 'moderate' and 'dubious' categories.

The classification of cases into conduct and neurotic disorders is also shown in Table 5 for cases from the screen positive group and control

Table 5. Clinical assessment of screen positive and screen negative cases by severity and type of disorder

Screen assessment	Type of disorder	Clinical assessment				
		Severity of disorder				
		Nil	Dubious	Moderate	Marked	Total
Not identified	Normal	42	0	0	0	42
	Conduct	0	4	2	0	6
	Neurotic	0	7	9	0	16
	Total not identified	42	11	11	0	64
Identified	Normal	1	0	0	0	1
	Conduct	0	14	44	23	81
	Neurotic	0	25	53	20	98
	Total identified	1	39	97	43	180

group respectively. In the former group, 81 conduct and 98 neurotic children were selected – a 1:1.2 ratio. In the control group, the false negatives divide into 6 conduct and 16 neurotic cases, a ratio of 1:2.7. Overall, the percentages of conduct and neurotic cases selected are 43.3 and 56.7 respectively, a ratio of 1:1.31. Among the cases identified by the screen, somewhat more of the conduct cases (23 cases or 28%) than of the neurotic cases (20 cases or 20%), fall in the 'severe' category, while slightly over half of each fall in the 'moderate' category. While the number of children classified as disturbed in the control group is small, there is an excess of neurotic cases, and a tendency for the neurotic cases to be rated as having more severe disturbance. Analysis of screen scores of children grouped by the conduct/neurotic classification, and by sex, will be provided in a future publication.

The relationship between severity of rated disturbance and screen weighted score is shown in Table 6. The proportion of cases in the 'moderate' and 'marked' categories tends to increase progressively as the weighted score increases. A converse progression is evident for the 'nil' and 'dubious' categories.

(d) Prevalence

On the basis of these data, one can estimate the total prevalence of disorder in the population studied. For this exercise, children with a 'dubious' degree of disorder are included with those considered to have no disturbance. Table 7 shows that estimates based on findings for the

Table 6. Distribution (%) of rated disorder in relation to weighted score

Clinical level of disturbance	Screen weighted score			
	Controls (N = 64)	Screen positives (N = 180)		
		0, 1, 2	3	4
Nil + dubious	82.8	32.5	18.6	4.2
Moderate + marked	17.2	67.5	81.3	95.8

rated cases indicate that 428 children in the population would show at least moderate disturbance, an overall percentage of 24.8.

(e) The efficiency of the screen

From the data in Table 5, the 'sensitivity' and 'specificity' of the screen can be calculated. Sensitivity refers to the screen's efficiency in identifying true positives, and is calculated by expressing 'true positives' as a percentage of all true cases found by psychiatric rating (Goldberg, 1972). The screen's 'specificity' or efficiency in detecting 'true negatives' is calculated by expressing the number of true negatives as a percentage of the total number of non-cases. These indices are, of course, a function of the definition of cut-off adopted for psychiatric disturbance. Taking 'dubious' cases as non-cases, the screen's sensitivity is 92.7% $\{140/(140+11)\} \times 100$, and specificity is 57% $\{53/(53+40)\} \times 100$. The overall efficiency of the screen in detecting valid positives and valid negatives is 79% $\{(140+53)/244\} \times 100$. If

Table 7. Total prevalence of psychiatric disorder

	No. in total population (a)	% of sample with psychiatric disorder (b)	Estimated number with disorder in total population (c = a × b)
Not selected on screening	1506	17.2	259
Selected on screening	217	77.8	169
Total	1723	—	428

$$\text{Total prevalence} = (428/1723) \times 100 = 24.8\%$$

'dubious' cases, on the other hand, are regarded as true cases rather than as non-cases, the respective figures become 89% and 95.5%, and overall efficiency becomes 90.2%.

Another crucial question is whether a statistically-derived weighting of our screen would prove more effective in discriminating between clinical positives and clinical negatives than the adopted system of allocating different weightings according to degree of deviancy on the screen measures employed. We tested this by dichotomizing the cases into those who proved to be disturbed and those not disturbed on clinical assessment. When using statistically-derived weightings from discriminant function analysis, the differentiation between true and false positives proved no more effective than the weighting scheme derived above.

(f) Behaviour rating scores of children identified by other criteria

How do teachers view children who are identified by sociometric or self-rating criteria? We can compare teacher ratings of the children falling above and below the cut-offs on these measures (i.e. children are dichotomized as scoring above or below the instrument cut-off, irrespective of whether they are screen positives or negatives). Table 8 shows that children falling above any sociometric criterion have significantly higher scores on the 3 Rutter criteria than those falling below the cut-off. In other words, children identified as having peer difficulties are also seen as deviant by teachers in terms of overall disturbance, antisocial and neurotic behaviour. On the other hand, children *not* identified by the JEPI Neuroticism cut-off tend to have higher Rutter scores, though the difference reaches significance only in the case of Rutter Antisocial score, i.e. children not identi-

fied by JEPI Neuroticism are seen by teachers as showing more antisocial behaviour.

(g) Neuroticism scores of children identified by other criteria

Repeating the above exercise with JEPI Neuroticism scores enables us to explore the relevance of neuroticism to deviance as indicated by teachers and peers (Table 9). Only in the case of Rejection does JEPI Neuroticism score differentiate between cases falling above and below the cut-off, with rejected children proving more neurotic ($P < 0.05$).

DISCUSSION

A number of important themes which merit more detailed discussion emerge from the above findings. These are (a) the efficiency of the screen, (b) the relationships between screen measures, (c) the neurotic-conduct ratio, and (d) prevalence.

(a) Efficiency of the screen

One of the most fundamental requirements in an intervention study such as the present one is that cases selected for inclusion in treatment regimes show genuine disturbance. In this respect, the multiple screen has been commendably sensitive, selecting only one false positive, and missing no cases showing marked disturbance. A lower cut-off may have captured some of the disturbed children missed by the screen (false negatives), but only at the risk of incorporating a higher percentage of non-disturbed children in treatment programmes. For example, a lower cut-off of 2 would have led to the inclusion of some 80 further cases, none of whom would have been rated as having a severe

Table 8. Comparison of Rutter B scores of children identified or not identified on other screen criteria

Screen criterion	Rutter Total score		Rutter Neurotic score		Rutter Antisocial score		No. of children	
	Identified	Not identified	Identified	Not identified	Identified	Not identified	Identified	Not identified
Isolation	9.81***	4.42	3.68***	1.27	3.37***	1.74	84	684
Rejection	10.18***	3.76	3.06***	1.16	4.08***	1.40	149	619
JEPI Neuroticism	4.06 NS	5.18	1.51 NS	1.53	1.30*	2.03	124	640

* $P < 0.05$; *** $P < 0.001$.

Table 9. Comparison of JEPI Neuroticism and Extroversion mean scores of children identified or not identified on other screen criteria

Screen criterion	JEPI Neuroticism		JEPI Extroversion†		No. of children	
	Identified	Not identified	Identified	Not identified	Identified	Not identified
Rutter Total	13.33 NS	14.09	16.35**	17.17	165	599
Rutter Neurotic	14.29 NS	13.85	15.81***	17.23	126	638
Rutter Antisocial	13.58 NS	14.02	16.70 NS	17.07	158	606
Isolation	14.90 NS	13.82	15.83***	17.14	83	683
Rejection	14.73*	13.75	16.44*	17.13	149	617

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

† Table 9 also shows the mean JEPI Extroversion scores for groups identified by teacher ratings. Children are significantly differentiated by extroversion scores on all comparisons, except that for Rutter Antisocial, with children above the cut-offs receiving more introverted scores, suggesting that, in terms of the kind of deviance measured by peer and teacher ratings, the introversion-extroversion continuum bears a more relevant and meaningful relationship than does neuroticism. The most significant findings ($P < 0.001$) are that children identified by peers as isolated and by teachers as neurotic prove to be more introverted in their self-ratings than those not so identified.

degree of disturbance and 74 of whom would show no disturbance. The adopted cut-off is believed to have been the optimum, and the subsequent discriminant function analysis tends to support this.

It is arguable that this efficiency is due in no small measure to a screen design that taps several domains of child functioning rather than one. This viewpoint gains some support from a comparison between the present multiple criterion screen and that of Rutter *et al.* (1975) employing the single criterion of teacher rating (Rutter Scale B2) in an Inner London Borough. Rutter *et al.* studied a population of 1689 10-year-old children, of whom 322 were selected as deviant by their screen. On the basis of more intensive investigation of a sample of these children and controls, including information from parental interview, a rating as to the presence and severity of psychiatric disorder was made. Of the 322 selected by the screen 139 proved to be disturbed and 183 not disturbed.

Table 10. Comparison of single and multiple criterion screening

(a) Single criterion screen (Rutter <i>et al.</i> 1975)			
	Disturbed	Not disturbed	Total
Selected	139	183	322
Not selected	289	1078	1367
Total	428	1261	1689

Kappa = 0.20 with s.e. 0.02.
Within the border totals, the maximum obtainable Kappa is 0.82.

(b) Newcastle multiple criterion screen			
	Disturbed	Not disturbed	Total
Selected	169	48	217
Not selected	259	1247	1506
Total	428	1295	1723

Kappa = 0.43 with s.e. 0.02.
Within the border totals, the maximum obtainable Kappa is 0.61.

Of the remaining 1367 not selected, they estimate that 289 were disturbed and 1078 not disturbed.

Our population was marginally greater (1723), and of these our multiple criterion screen selected 217 as deviant, of whom 169 were estimated to be at least moderately disturbed clinically. Hence, for populations similar both in size and estimated rate of disturbance, the multiple criterion screen selects a smaller number of deviant cases, of whom 77.8% prove to be disturbed clinically, and the single criterion screen selects a larger number of cases, of whom 43.3% prove to be disturbed clinically. One way of comparing the efficiency of these 2 screens is by studying the Kappas (roughly equivalent to an efficiency correlation) by using the data in Table 10. The single criterion screen achieves a Kappa of 0.20 (with a maximum obtainable Kappa of 0.82) and the multiple criterion screen achieves a Kappa of 0.43 (with a maximum obtainable Kappa of 0.61). These results suggest that a multiple screen is preferable in identifying cases for an intervention programme and that reliance on a single criterion screen may be penalized by excessive false positive rates.

To adopt the single measure of teacher rating, as has been done in many studies, as a means of reducing the burden of screening, may be in more than one sense false economy: our experience has shown that the gathering of teacher ratings is a much more time-consuming and tiresome exercise than is the collection of self-rating and sociometric data on classroom groups. The punitive false positive rate when employing teacher ratings as the sole index has already been pointed out by Tizard (1968) and Booth & Taylor (1973), in discussion of the Rutter scale. Rutter himself acknowledges the corrective role of multiple screen measures (Rutter, 1977), and the strategy and results of the present study would seem to reinforce this, while also underlining the important contribution that the Rutter scale can make.

The situation specificity of disturbed behaviour (Mischel, 1968; Kolvin *et al.* 1977) makes it inevitable that a school-based screen will miss children who show disturbance predominantly or exclusively in the home. This omission is not of great importance in the present study, since our major concern was the identification and treatment of disturbance in school. However, we shall be reporting elsewhere (Nicol *et al.* 1980) on the relationships between our school screen data and data obtained in the home, which

will cast some light on the issue of generality of disturbance. It is interesting to note that the majority of disturbed children who are missed by the screen are classified as neurotic on subsequent psychiatric assessment.

(b) Relationships between screen measures

Where a number of criteria are employed in a screen, economy must also be considered, in conjunction with efficiency. In other words, is the overlap between criteria such that certain of them are redundant? The Rutter Antisocial criterion has high overlap with the Rutter Total score and correlates highly with it, suggesting that it may have little independent value. The high specificity of JEPI Neuroticism and a pattern of largely non-significant correlations with other criteria can be interpreted in 2 ways. First, the neuroticism dimension may not validly identify disturbance; or, secondly, it may identify some of the disturbed children who are missed when only teacher and peer screen measures are employed. The psychiatric ratings do not support the former interpretation of the 16 cases selected solely by JEPI Neuroticism: all were considered to show some degree of disturbance, and their distribution across 'dubious', 'moderate' and 'severe' categories was 4, 9 and 3 respectively. The self-rating selection criterion would therefore appear to be a valuable adjunct to teacher and peer information.

The lack of correlation between Rutter scores and neuroticism on Eysenck questionnaires has also been observed by Powell (1977), with psychoticism on the Junior EPQ being the only dimension to relate significantly to Rutter Total. JEPI Neuroticism scores might be expected to relate most closely to the Rutter Neurotic scores. No relationship is observed in the present study, and closer inspection of the degree of overlap in the two scales may suggest one explanation. Only 4 of the 9 Rutter items employed in the extended neurotic subscale are explored to any extent in the JEPI, and 2 are emphasized more than the others - 'worrying' and 'being miserable/unhappy'. The content of the Rutter Neurotic items - K (twitches, mannerisms), N (absenteeism for trivial reasons), Q (disobedience), R (fussy, over-particular) and W (tears on arrival at school) - is not represented in the JEPI Neuroticism scale. The lack of

correlation may therefore be due to some extent to the fact that the scales have little content in common. On the other hand, in their presentation of data showing a relative lack of agreement between JEPI responses and teachers' ratings of emotional stability, perseverance, sociability and impulsiveness, Eysenck & Cookson (1969) comment that even obtaining ratings on what seem to be semantically identical variables gives no guarantee of agreement. As has been suggested earlier, in relation to JEPI Neuroticism and Extroversion scores of children identified by the Rutter and sociometric criteria, the introversion-extroversion dimension may have a closer relationship than neuroticism to the type of disturbance measured by these teacher and peer ratings. This is also borne out by correlative data and will be described in a later report.

The pattern of relationships between screen instruments that is observed is suggested by Wilson (1974) to be typical of assessments of disturbance which draw on multiple sources, i.e. a closer correspondence between teacher ratings and sociometric ratings than between either of these and self-ratings. Semler (1960) suggests that the correlation between teacher and peer ratings indicates a common 'social reaction' and that teachers are conscious of the role of social acceptability in adjustment. This correlation may also reflect situation specificity to some extent. Teacher and peer ratings can be seen as being, to a large extent, a function of the child's behaviour in the school setting, while self-ratings are relatively less bound to one setting and can span a variety of situations. It may be therefore that the ambiguity of minimal overlap between disturbance as assessed in the school and in the home is not necessarily resolved by confining the assessment to one situation. The nature of the measuring instruments employed, and their precise situational referents may have much to do with the degree of consistency and overlap that is obtained. Another possibility worthy of investigation is that consistency and overlap may be a function of *degree* of disturbance, being more marked with greater disturbance than with disturbance of minimal extent.

Examining the sociometric indices more closely, it is evident that the Isolation criterion has a small independent contribution (14%) and shows lower correlations with behavioural criteria than does Rejection. While this limited

specific contribution may constitute an argument against the worth of its inclusion, the extent of its correlation with Rejection indicates that these two aspects of malfunction *vis-à-vis* peers are not identical. We have discussed elsewhere data obtained on younger children which suggest that Isolation is less closely related than Rejection to disturbed behaviour, whether this is rated by parents or teachers (Kolvin *et al.* 1976; Macmillan *et al.* 1978). Lack of popularity may not be synonymous with poor adjustment, and the lower test-retest reliability of Isolation suggests that it is generally less stable and enduring than Rejection. The independent validity of Isolation in the present population requires further analysis. A score above the Isolation cut-off was in itself not sufficient for inclusion of a case as a screen positive, so the type of analysis we have given for severity of disturbance ratings in cases selected by JEPI Neuroticism alone is not possible in the case of the Isolation criterion. However, despite the relatively low independent contribution of the two sociometric indices, it is worth noting that, of the 98 children who were identified by the *minimum* total screen score of three points, 18 were identified with a contribution from the Isolation criterion, and 33 with a contribution from Rejection. In other words, had the sociometric criteria not been employed, some 51 cases would have been missed.

(c) The neurotic-conduct ratio

The overall ratio of 1.31:1 neurotic to conduct disorder cases is one of the most intriguing findings in the study. Such a ratio is somewhat surprising in an industrialized urban population, especially since 'mixed' cases are combined with cases in the conduct category. It is also inconsistent with traditional findings in the literature (e.g. Rutter, 1967; Rutter *et al.* 1970), where a preponderance of conduct disorders is usually reported. It must be borne in mind in this connection that the self-rating criterion (JEPI Neuroticism) employed in decisions on classification in the present study had an input specific for neuroticism. In other words, the multiple screen criteria used may have been less weighted towards ascertainment of conduct disturbance than the criteria employed in other studies in which self-rating questionnaires yielding information on emotional disturbance have not been

employed. The estimated ratio of types of disorder for a population could, to some degree, reflect the bias in the measuring instruments employed towards one type of disorder than another. Bias in raters is also possible, and in the context of school-based screening much discussion has centred around whether or not teachers overemphasize acting out, antisocial disorder (Wickman, 1928; Walsh & O'Connor, 1968; Fremont *et al.* 1976).

(d) Prevalence

The overall estimated prevalence rate of 24.8% observed here is of interest from several viewpoints. First, there is a remarkable correspondence with the figure quoted by Rutter *et al.* (1975) from screening conducted in an Inner London Borough (25.4%). While the equivalence of disturbance rate in these 2 geographical areas is not thereby established, it is at least clear that both areas yield substantially higher rates than that observed in rural populations such as the Isle of Wight (Rutter *et al.* 1970). This invites further research into specific mediating factors. Secondly, the high rate of disturbance in children both in secondary and primary schools (Kolvin *et al.* 1977) constitutes ample justification for exploring alternative ways of dealing with such problems within the community.

This research was funded by the Department of Education and Science.

REFERENCES

- Booth, G. & Taylor, M. (1973). Some problems in the rating of child behaviour by teachers. *Journal of Association of Educational Psychologists* 3, 13-17.
- Bower, E. (1969). *Early Identification of Emotionally Handicapped Children in School*. Thomas: Springfield, Ill.
- Brophy, J. & Good, T. (1970). Teachers' communication of differential expectations for children's classroom performance: some behaviour data. *Journal of Educational Psychology* 61, 365-374.
- Davies, B. (1976). Piggies in the middle or 'Who, sir? No, not me, sir'. In *The Disruptive Pupil in the Secondary School* (ed. C. Jones-Davies and R. G. Cave), pp. 26-34. Ward Lock Educational: London.
- Eysenck, H. J. & Cookson, D. (1969). Personality in primary school children. 2. Teachers' ratings. *British Journal of Educational Psychology* 39, 123-130.
- Eysenck, S. B. G. (1965). *Manual of the Junior Eysenck Personality Inventory*. University of London Press: London.
- Fremont, T. S., Klingsporn, M. J. & Wilson, J. H. (1976). Identifying emotional disturbance in children - the professionals differ. *Journal of School Psychology* 14, 275-282.
- Garner, J. & Bing, M. (1973). The elusiveness of Pygmalion and differences in teacher-pupil contacts. *Interchange* 4, 34-42.
- Goldberg, D. P. (1972). *The Detection of Psychiatric Illness by Questionnaire*. Maudsley Monograph no. 21. Oxford University Press: Oxford.
- Kolvin, I., Garside, R. F., Nicol, A. R., Macmillan, A. & Wolstenholme, F. (1976). Maladjusted pupils in ordinary schools. *Special Education: Forward Trends* 3, 15-19.
- Kolvin, I., Garside, R. F., Nicol, A. R., Leitch, I. M. & Macmillan, A. (1977). Screening school children for high risk of emotional and educational disorder. *British Journal of Psychiatry* 131, 192-206.
- Macmillan, A., Walker, L., Garside, R. F., Kolvin, I., Leitch, I. M. & Nicol, A. R. (1978). The development and application of sociometric techniques for the identification of isolated and rejected children. *Journal of Association of Workers for Maladjusted Children* 6, 58-74.
- Mischel, W. (1968). *Personality and Assessment*. John Wiley: London.
- Mitchell, S. & Shepherd, M. (1966). A comparative study of children's behaviour at home and at school. *British Journal of Educational Psychology* 36, 248-254.
- Neligan, G. A., Prudham, D. & Steiner, H. (1975). *The Formative Years: Birth, Family and Development in Newcastle-upon-Tyne*. Oxford University Press for the Nuffield Provincial Hospitals Trust: Oxford.
- Nicol, A. R., Kolvin, I., Garside, R. F., Macmillan, A., Wolstenholme, F. & Leitch, I. M. (1980). A multiple criterion screen for identifying secondary school children with psychiatric disorder. II. Studies in validity (in preparation).
- Powell, G. E. (1977). Psychoticism and social deviancy in children. *Advances in Behaviour Research and Therapy* 1, 27-56.
- 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. (1977). Surveys to answer questions: some methodological considerations. In *Epidemiological Approaches in Child Psychiatry* (ed. P. J. Graham), pp. 1-30. Academic Press: London.
- Rutter, M., Tizard, J. & Whitmore, K. (1970). *Education, Health and Behaviour*. Longman: London.
- Rutter, M., Cox, A., Tupling, C., Berger, M. & Yule, W. (1975). Attainment and adjustment in two geographical areas. 1. The prevalence of psychiatric disorder. *British Journal of Psychiatry* 126, 493-509.
- Semler, I. J. (1960). Relationships among several measures of pupil adjustment. *Journal of Educational Psychology* 51, 60-64.
- Tizard, J. (1968). Questionnaire measures of maladjustment: a postscript to the symposium. *British Journal of Educational Psychology* 38, 9-13.
- Walsh, J. F. & O'Connor, J. D. (1968). When are children disturbed? *Elementary School Journal* 68, 353-356.
- Wickman, E. K. (1928). *Children's Behaviour and Teachers' Attitudes*. Commonwealth Fund: New York.
- Wilson, J. A. (1974). Adjustment in the classroom. 2. Patterns of adaptation. *Research in Education* 11, 17-29.
- Wrate, R. M., Nicol, A. R. & Kolvin, I. (1980). Reliability of clinical ratings of psychiatric disorder from standard case records (in preparation).