

**Part Two**  
**Method and background**

### 3 Aims and method

#### Summary

As mentioned in our introductory chapter, the main aims of our research project were to identify maladjusted children in ordinary schools and to evaluate the effectiveness of different treatment approaches applied to them. In comparing these approaches, seven hypotheses were tested. We describe these below (p. 36) and outline how we tested them.

A total of almost 600 children were selected to take part in the study by a series of screening methods (involving approximately 4300 children) that identified those with signs of disturbance. These methods detected junior schoolchildren, aged seven to eight years, who showed some signs of having social or psychiatric disturbance or learning problems. These we have described as being 'at-risk' children. With the senior schoolchildren, aged eleven to twelve years, only those who showed relatively clear-cut psychiatric disturbance were included. The screening tests, then, differed for junior and senior children but, basically, they relied on a classroom multiple criterion screen.

Subsequently, additional information was gathered on those identified, from the teacher, the child, and his or her parents. The data were used to place each child's disturbance into the category of either conduct or neurotic disorder, and to rate the degree of severity of the problem.

The children thus selected as being 'screen-positive' were randomly allocated to various treatment or control regimes. Each regime was mounted in each of the six junior schools; a similar plan was adopted for the senior children. To ease the impact on the schools, different

## 36 Help Starts Here

regimes were studied in two consecutive years. We summarize the elements of the treatment programmes, and outline how the effects of these treatments were evaluated, taking into account various initial differences between the groups. We also touch on the problem of missing data, which was not, however, found to affect the conclusions drawn.

### Aims and hypotheses

In aiming to identify maladjusted schoolchildren and compare ways of treating them we mounted two programmes of treatment – one for younger children (aged seven to eight years) and one for older children (aged eleven to twelve years). Each programme consisted of four regimes – three treatment regimes and one no-treatment regime (the controls). The regimes for both programmes have been outlined in Chapter 1 and are summarized in *Table 3(1)*.

Table 3(1) *Basic design of research project*

| school                 | type of therapy and year of treatment                   |  |  |   |
|------------------------|---|--|--|---|
|                        | at-risk controls (ARC)                                  | parent counselling-teacher consultation (JPC)                              | group therapy/playgroups (PG)                                  | nurture work (NW)   |
| junior (total n = 270) | (project year 1)<br>n = 67                              | (project year 1)<br>n = 69   | (project year 2)<br>n = 74                                     | (project year 2)<br>n = 60                                |
| senior (total n = 322) | maladjusted controls (MC)<br>(project year 1)<br>n = 92 | parent counselling-teacher consultation (PC)<br>(project year 1)<br>n = 83 | group therapy/senior groups (SG)<br>(project year 2)<br>n = 73 | behaviour modification (BM)<br>(project year 2)<br>n = 74 |

*Note:* both the junior and senior school programmes took place within six schools.

The precise aims of the research were to test the following hypotheses, which applied both to the younger and the older children.

- (1) The four regimes (which include the no-treatment regime) differ in effectiveness in reducing maladjustment; that is, some forms of management are better than others.
- (2) One or more of the three treatment regimes is more effective than the no-treatment regime (the controls) in reducing maladjustment; that is, any treatment is better than no treatment at all.
- (3) Regimes differ in effectiveness according to the diagnostic category into which the child falls; that is, some regimes are more effective in helping children with neurotic disorders than those

with conduct disorders, and vice versa. This hypothesis relates to mutually exclusive categories of children, whereas hypothesis (5) refers to different patterns of behaviour which may co-exist in the same child.

- (4) Regimes differ in effectiveness according to the sex of the child; that is, some regimes are more effective for boys, others for girls.
- (5) Regimes differ in effectiveness according to patterns of behaviour co-existing in any one child; that is, some regimes are more effective in reducing the neurotic component of a child's behaviour and others in reducing his or her antisocial behaviour, so that a child's neurotic behaviour may be reduced while his or her antisocial behaviour remains unchanged.
- (6) Irrespective of treatment regime, improvement is related to diagnostic category; that is, children with conduct disorders differ from those with neurotic disorders in relation to degree of improvement.
- (7) Irrespective of treatment regime, improvement is related to the sex of the child; that is, boys differ from girls in relation to degree of improvement.

#### Method of testing hypotheses

Our first task was to detect suitable cases, which we did by screening, as subsequently described. We then assessed and classified these screen-positive children and randomly allocated them by school class into one of four regimes, ensuring that every regime was represented in each school. Assessments of the children were undertaken at the baseline and on two and three subsequent occasions for the juniors and seniors respectively (see *Table 3(5)*). Data analysis involved comparisons of the effects of the four regimes.

#### TIMING

At the outset we were aware that the project would make heavy demands on school staff over and above their normal duties so, to lessen the impact, intervention was spread over two consecutive years, involving children of the same age and general characteristics in both years. During the first year the control regime and one specific treatment regime (parent counselling-teacher consultation) were studied in both the junior and senior schools. The remaining treatments (nurture work and group therapy for the juniors and behaviour modification and group therapy for the seniors) were studied in the second year (see *Table 3(1)* and *Figs 3(1)* and *3(2)*). It was not practicable to randomize the treatment methods within both years

## 38 Help Starts Here

of the study for two main reasons – first, insufficiency in numbers of screen-positive cases in a single year and, second, insufficiency of resources to carry out all treatments simultaneously.

### Screening – general considerations

Several basic principles guided the design of our screening method with both junior and senior schoolchildren. First, because we were interested primarily in intervention in school settings we sought screen measures that could be applied in the schools. Had we been interested in, for example, prevalence of psychiatric disturbance, this would have demanded a different type of screening, including the home. Second, as we felt that as little energy as possible should be deflected from the main object of the study – getting special help to the children – we wanted the screen to be conducted as rapidly and economically as possible. Disruption of school routine had to be minimized, particularly because of the value of ensuring staff co-operation in a follow-up project of this nature. Third, as we wanted to explore the many ways in which disturbance may manifest itself within the school, we considered it preferable to develop a screen battery that would draw upon three sources of information – teachers, peers, and the children themselves.

The schools in which we worked – six junior schools and six junior high or comprehensive schools (the senior schools) – were broadly representative of those in the Newcastle upon Tyne and Gateshead area; the social class distribution of the area is slightly below the national average (Neligan, Prudham, and Steiner 1974).

Screening was conducted with children entering the first year of junior or senior schooling in the academic year beginning September, 1972, and then again one year later with the new intake (see p. 37).

### Screening – junior schoolchildren

#### THE POPULATION

Taking the two cohorts together, approximately 1000 children were screened. A description and analysis of the screening method are provided elsewhere (Kolvin *et al.* 1977). The mean age of this population of children was seven years and nine months and the sex distribution was about equal (52 per cent were boys and 48 per cent girls).

## DESIGNING THE MULTIPLE CRITERION SCREEN

A child must make three major adjustments at school. First, there are the formal educational and academic demands; second, there must be accommodation to requirements for behavioural control; and, third, appropriate social relationships with peers have to be established. We proposed that signs of failure, whether mild or severe, in any one of these areas may have repercussions in any of the others: also, mild disorders may progress to more serious behavioural or educational problems over time. We therefore described these children as being 'at risk'. With these points in mind, we felt that for our purposes a multiscreen model (e.g. that of Bower, 1960-69) was preferable to one that relied on a single measure. We scanned the literature for screen measures that could be applied in schools and were likely to be reliable, valid, and reasonably efficient predictors of disorder.

We eventually decided to use five screen criteria, which were as follows:

- (i) and (ii) The sociometric criteria of *isolation* and *rejection*. Each child in a class was asked to choose three classmates they would like to sit beside in class, and three they would like to play with at play-time; in addition, they were asked which they would *not* like to sit beside or play with. This procedure yielded two scores: isolation, which is defined as a lack of positive choice, and rejection, which is defined as receipt of a large number of negative choices. In a class of approximately thirty children, a child was considered to be isolated if he or she scored nought, or one positive choice, and to be rejected if he or she received fourteen or more negative choices. Each of these cut-off scores was intended to identify about 6 per cent of the population, judged by the pilot study data, but in practice we found that the yield was higher than 6 per cent; in other words, there were more isolated and rejected children than we expected. The sociometric criteria are described more fully elsewhere (Macmillan *et al.* 1978). Test-re-test reliability was assessed for two age levels. First, administration on two occasions (four-and-a-half weeks apart and involving a total of fifty-eight seven-year-old children in two primary classes) yielded a correlation of 0.64 for isolation and 0.87 for rejection. A similar exercise with sixty-one eleven- to twelve-year-olds yielded correlations of 0.72 for isolation and 0.87 for rejection.
- (iii) *Reading*. In the Isle of Wight study (Rutter, Tizard, and Whitmore 1970), the definition of educational backwardness was reading accuracy or comprehension twenty months or more below the

## 40 Help Starts Here

child's chronological age, and this included 7.9 per cent of the population. Some of the children in our study (i.e. those who were just seven years old) were too young for a similar definition to be used with any degree of confidence, so we selected a cut-off of a reading quotient (RQ) of seventy-five or less on the Young (1968) group reading test. This meant that, at seven to eight years of age, the selected children were non-readers on this test. We appreciated that this would produce a higher yield than the Isle of Wight study. In fact, it led to an inclusion of 12.2 per cent of the population.

The Young Group Reading Test was developed to provide a group measure of reading ability that permitted easy application and quick marking. Two parallel forms are available, each comprising forty-five items. The first fifteen of these require the child to select which one of between three and five words matches a given picture. With the remaining thirty items the task is to identify synonyms in a multiple-choice sentence-completion format. The two sections are timed, with four and nine minutes respectively being allowed. Young's standardization sample consisted of 7400 children in an urban area, aged six years and six months to twelve years and eleven months. However, because 50 per cent of the scores of children over ten years were above the test ceiling, the number of results making an effective contribution to the tables of norms was about 5600.

Young reported satisfactory reliability of 0.95. Validity data showed correlations of 0.88 with Neale's Analysis of Reading Ability (A) accuracy score, 0.88 with Vernon's Graded Word Reading Test, and 0.88 with the National Foundation for Educational Research (NFER) Sentence Reading Test (1) ( $n = 80$  in all cases).

- (iv) *Behaviour – the Rutter teacher scale B2.* This is a well-known and established scale completed by teachers regarding children's behaviour in school. It yields a total, a neurotic, and an antisocial score. Rutter (1967) and Rutter, Tizard, and Whitmore (1970) found that a cut-off of nine or more on the total score had discriminative value. In their Isle of Wight study it selected about 10 per cent of the boys and 4 per cent of the girls – an average 7 per cent. This cut-off produced nearly 30 per cent of children in our pilot study in Newcastle upon Tyne, and over 20 per cent of the children studied overall. We decided, therefore, to use a slightly more rigorous criterion, and raised the cut-off to ten. At this level 17 per cent of the population were included, more than double the rate reported by Rutter (Kolvin *et al.* 1977).

- (v) *Absenteeism*. This is item 'N' on the Rutter B2 scale – 'tends to be absent from school for trivial reasons'. Children were selected on this criterion if the rating 'certainly applies' was made.

#### IDENTIFICATION OF 'AT-RISK' CHILDREN

Identification by any one or more criteria was taken as indicating that the child might be 'at risk'.

### Screening – senior schoolchildren

#### THE POPULATION

Taking the two consecutive years of screening together, approximately 3300 children were screened. The sex ratio was about equal and the mean age was eleven years and eight months. Again, a more detailed description and analysis of the screening methods are provided elsewhere (Macmillan *et al.* 1980).

#### DESIGNING THE MULTIPLE CRITERION SCREEN

With the senior children we sought measures that reflected the perceptions of the child, his or her peers, and the teacher. We felt that the most commonly used screening technique – teachers' ratings – might be inadequate if teachers' views were not supplemented by other information. For example, teachers may overlook quiet, passive, but potentially disturbed children (Garner and Bing 1973) and may not be sufficiently attuned to the interpersonal difficulties that some children may be experiencing. We thought that sociometric data might supply more accurate information on social functioning and that self-ratings, in complementing teacher- and peer-derived data, might reveal the personal unease and concerns that both other sources of information may bypass.

We employed, then, three screen measures yielding six criteria contributing to identification:

- (i) *Teachers' ratings*. The scale used here was the same as that employed with the younger children – the Rutter B2 scale. As mentioned previously, as well as a total score it yields two sub-scale scores, one for neurotic, the other for antisocial behaviour. To increase the reliability of the two sub-scales the number of items contributing to them was enlarged in the present study; this was following an inspection of Rutter's data (1967), bearing on discrimination of the two types of disorder. Items were added on the basis that they (a) differentiated the psychiatric group from



## 42 Help Starts Here

the controls and (b) differentiated the diagnostic groups. The revised sub-scales therefore consisted of the following items concerning neurotic behaviour: G, H, J, K, N, Q, R, V, and W; and concerning antisocial behaviour: A, B, D, E, O, P, S, T, and Z. Details are provided elsewhere (Macmillan *et al.*, 1980).

- (ii) *Sociometry*. The sociometric instrument was the same as that employed with the junior children, but it was used in a slightly different way, as explained in the following section on cut-offs.
- (iii) *Self-rating*. The measure employed here was the Junior Eysenck Personality Inventory (JEPI) (Eysenck 1965). This is a sixty-item questionnaire yielding an extroversion-introversion score, a lie score, and a neuroticism score. It was with the last of these dimensions that we were particularly concerned. The high scorer on neuroticism is likely to be 'moody, touchy, anxious, restless, rigid' (Eysenck 1965:3). These are the characteristics associated with instability; the stable person (a low scorer on neuroticism) is likely to be calm, carefree, easy-going, and reliable.

### ESTABLISHING CUT-OFFS AND ASSIGNING WEIGHTINGS

It was decided to take extreme scores, on each of the screen measures, as indicators of maladjustment (Macmillan *et al.* 1980). The actual scores used as cut-offs were decided in most cases by examining published data on the characteristics of the instrument. For the sub-scales 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, there is the pitfall that it is theoretically possible for a child to, obtain a very deviant score on one specific measure only, but to, nevertheless, be excluded because his or her *summed* weighted score is not sufficiently high. To avoid this we adopted a weighting system that 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 nine has been regarded as providing the best discrimination between children attending child guidance clinics and a normal sample (Rutter, Tizard, and Whitmore 1970). We retained this cut-off, assigning it a weighting of one point towards the deviance classification. In addition, we gave the more extreme score of fifteen or over a weighting of two. This was arrived at by adding one standard deviation to Rutter's original cut-off. A cut-off of nine identified about 12 per cent of the sample, while a

cut-off  
from  
Cu  
whic  
four  
cent  
shou  
ing s  
who  
weig  
score  
aspe  
W  
guid  
so th  
were  
selec  
rejec  
and  
were  
C  
mea  
aver  
twen  
scor  
with  
thos  
poir  
Th  
in T  
Tabl  
—  
sourc  
infor  
—  
peer  
teach  
chik  
—  
T  
sun

cut-off of fifteen identified 2 per cent. These cut-offs were different from those used for the junior schoolchildren.

Cut-offs were assigned on the neurotic and antisocial sub-scales which gave rise to a yield closest to that of the total score. A score of four on the neurotic and antisocial sub-scales identified 11 and 14 per cent respectively, and hence was assigned a weighting of one. We should emphasize that our use of the sub-scales as providing weighting scores was different from the use by Rutter, Tizard, and Whitmore who used the two sub-scales for diagnostic purposes. The three Rutter weighted scores were added together to contribute to the total screen score. The use of the two sub-scales ensured that important specific aspects of behaviour were taken into consideration.

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 picked out 14 per cent of the pilot children. For rejection, twelve or more negative choices were taken as the cut-off and this also selected 14 per cent. Scores on or beyond these cut-offs were each weighted one point.

Cut-offs of one-and-a-half and two standard deviations above the mean for neuroticism were taken for the JEPI, with the scores being averaged so as to be equivalent for both sexes. These scores were twenty and twenty-three respectively: 17 per cent of the pilot sample scored one-and-a-half standard deviations above the mean. Children with scores of twenty to twenty-two were allotted two points, and those with more extreme scores of twenty-three or twenty-four, three points.

The screen tests used for both junior and senior children are shown in *Table 3(2)*.

*Table 3(2) Screen tests used in Newcastle upon Tyne Action Research Project*

| <i>source of information</i> | <i>junior school</i>                                | <i>senior school</i>                       |
|------------------------------|---|--|
| peers                        | sociometry  | sociometry                                 |
| teacher                      | Rutter teacher scale B2<br>absenteeism              | Rutter teacher scale B2                    |
| child                        | reading assessment<br>(Young group<br>reading test) | Junior Eysenck<br>Personality<br>Inventory |

The children's scores on each of the screen measures were summed. Those obtaining a total of three or more points were

## 44 Help Starts Here

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 3(3)* 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 by various combinations of scores from the three instruments (sociometric, JEPI, or Rutter B2 scale). The maximum weighted score that could be obtained was nine. The weighted score could therefore yield information concerning not only the presence or absence of disturbance, but also its level of severity.

Table 3(3) *Weighting system used to identify children in senior schools*

| <i>instrument</i>                          | <i>scores used</i>                                  |            | <i>weighting</i> |
|--|---|------------|------------------|
| Rutter teacher scale B2                    | total score   | 9-14       | 1                |
|  |   | 15 or more | 2                |
|  | neurotic sub-scale<br>(Newcastle<br>modification)   | 4 or more  | 1                |
|  | antisocial sub-scale<br>(Newcastle<br>modification) | 4 or more  | 1                |
| sociometry                                 | isolation   | 0 or 1     | 1                |
|  | rejection   | 12 or more | 1                |
| Junior Eysenck<br>Personality<br>Inventory | neuroticism   | 20-2       | 2                |
|  |   | 23-4       | 3                |

*Note:* children with a weighting score of three or more were regarded as screen positive.

### Additional data (see Appendix 1)

The baseline data provided by the screening techniques were supplemented by information obtained from interviewing the parents (on family and social conditions and the child's behaviour), from completion, by the teacher, of more detailed classroom behaviour checklists, and from some group and individual psychological tests (*Table 3(4)*). For screen purposes the Young Reading Test was used, with the Holborn Test used as an additional baseline measure. For all subsequent assessments the Holborn Test was used. The correlation between these tests was 0.9. All the data amassed were then studied by a psychiatrist in order to arrive at a diagnosis (conduct or neurotic disorder) on each child and also to rate them according to the degree of severity of the problem. In a very small number of cases it was not possible to arrive at such a clear-cut diagnosis.

Rutter, Tizard, and Whitmore (1970) managed to classify 90 per

Table 3(4) *Additional assessments used*

| <i>source of information</i> | <i>junior school</i>  | <i>senior school</i>  |
|------------------------------|---|---|
| child                        | reading<br>verbal and non-verbal<br>IQ                                | ability tests<br>school attitude<br>questionnaire                     |
| teacher                      | classroom behaviour<br>scale  | classroom behaviour<br>scale  |
| parent                       | social data<br>parent attitudes<br>child behaviour and<br>temperament | social data<br>parent attitudes<br>child behaviour and<br>temperament |
| clinical<br>staff            | clinical assessment   | clinical assessment   |

cent of their cases into conduct, neurotic, and mixed categories. However, they reported that their mixed group had much in common with their pure, conduct group. We therefore decided to split our cases into neurotic and conduct disorders by combining the conduct and mixed groups, thereby forming an expanded conduct-disorder group. Previously we had found that there was high agreement between clinicians in such a classification ( $\kappa = 0.9$ ). In addition, the clinicians rated all cases according to the extent of disturbed behaviour on four-point scales, namely: 1 = nil; 2 = dubious; 3 = moderately severe; and 4 = markedly severe. For this research's purposes three scales of disturbed behaviour were employed: *overall severity*, *antisocial* behaviour, and *neurotic* behaviour. The sources of information upon which the ratings were based were the behaviour and temperament scales derived from parental interviews and the behavioural information available from assessments in the school.

#### **Allocation to treatment or control groups**

The children who had been selected according to the criteria outlined above, i.e. the screen-positive children, were allocated by school class, at random, to various treatment or control regimes. The use of untreated controls for evaluation of treatment effects may be ethically justified where resources are inadequate to meet the very basic needs of the community, when it is acceptable to allocate randomly rather than to allow selective factors to determine which children are helped. Controls are certainly justified where there are doubts about treatment effectiveness; in fact it is from such controlled studies that we obtain evidence not only of ineffectiveness but also of the possible

## 46 Help Starts Here

Table 3(5) *Timing of assessments after baseline assessment (in months)*

| <i>treatment regime</i>                     | <i>end of treatment: limited assessment of seniors only</i> | <i>midline follow-up: senior and junior programmes (approximate)</i> | <i>final follow-up: senior and junior programmes</i> |
|---|---|--|--|
| <i>juniors</i>                              |   |  |  |
| at-risk controls                            | —   | 18–22  | 36   |
| parent counselling-<br>teacher consultation | —   | 18–22  | 36   |
| nurture work                                | —   | 17–19  | 36   |
| group therapy/<br>playgroups                | —   | 18–20  | 36   |
| <i>seniors</i>                              |   |  |  |
| maladjusted controls                        | 15–18   | 15–18  | 36   |
| parent counselling-<br>teacher consultation | 15–18   | 15–18  | 36   |
| behaviour modification                      | 7   | 15–18  | 36   |
| group therapy                               | 7   | 15–18  | 36   |

*Note:* for the majority of cases the midline follow-ups were completed over a narrower band of time than was the case with the other assessments.

adverse effects of certain forms of psychotherapy (California Youth Authority 1970).

There were two types of controls in operation: first, control groups in the schools in which treatment programmes were being undertaken (within-school controls) and, second, controls in schools where no treatment was being given (between-school controls). Subsequently, we discovered that the differences between schools were such that it would have been inappropriate to rely on between-school controls and, therefore, data relating to these are not presented in this book. One problem of having the control pupils in the same school as the treated pupils was the possibility that the beneficial effects of treatment may have spread to the former. While we were aware of this possible contamination of the within-school controls, it was noted that this contamination would tend to *reduce* differences between treated and control regimes, rather than exaggerate them.

Within certain practical constraints, we tried to ensure that the schools used were reasonably representative of state-run schools in the cities of Newcastle upon Tyne and Gateshead. These cities are fairly typical of the large, industrialized conurbations in the north of England, with their attendant economic and social problems. Indeed, this area has traditionally been associated with severe economic difficulties, and we believe that this is one of the major reasons for the relatively small-scale influx of immigrants.

Figure 3(1) Flow chart of study: junior school programme (this programme took place in each of six junior schools)

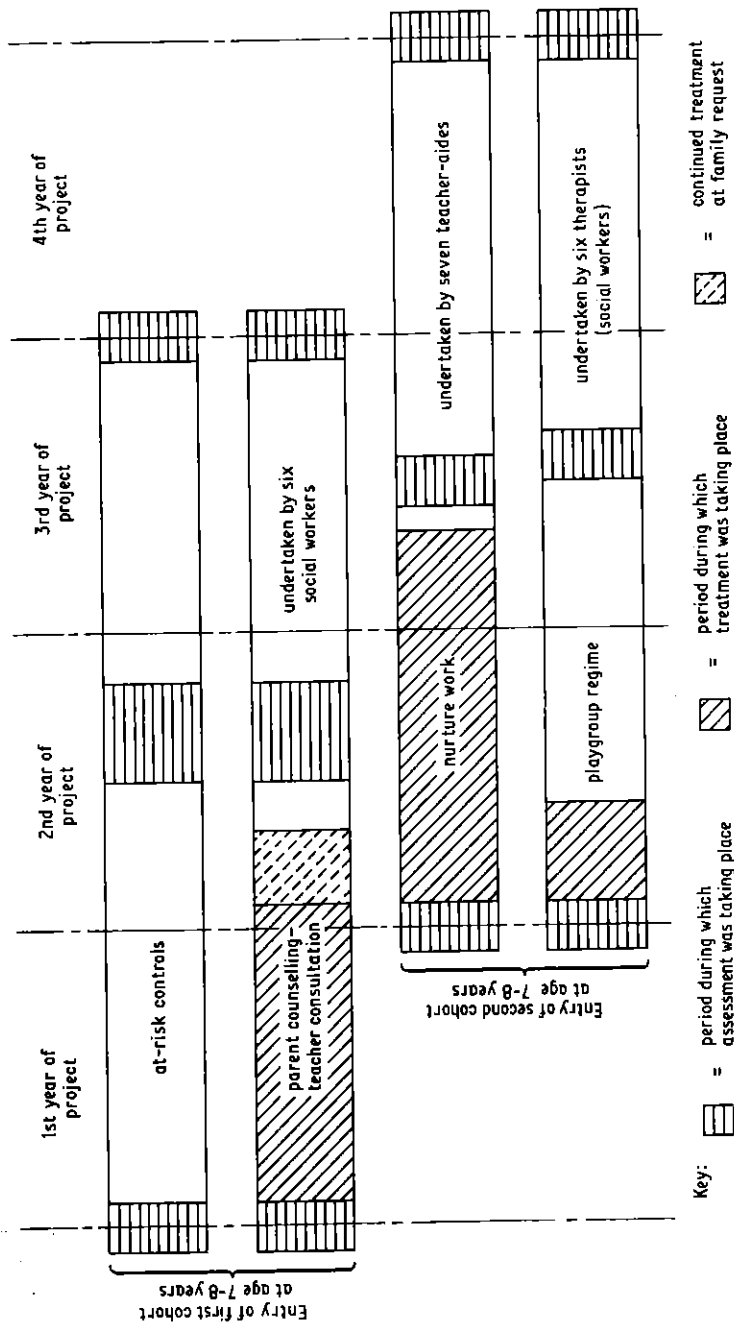
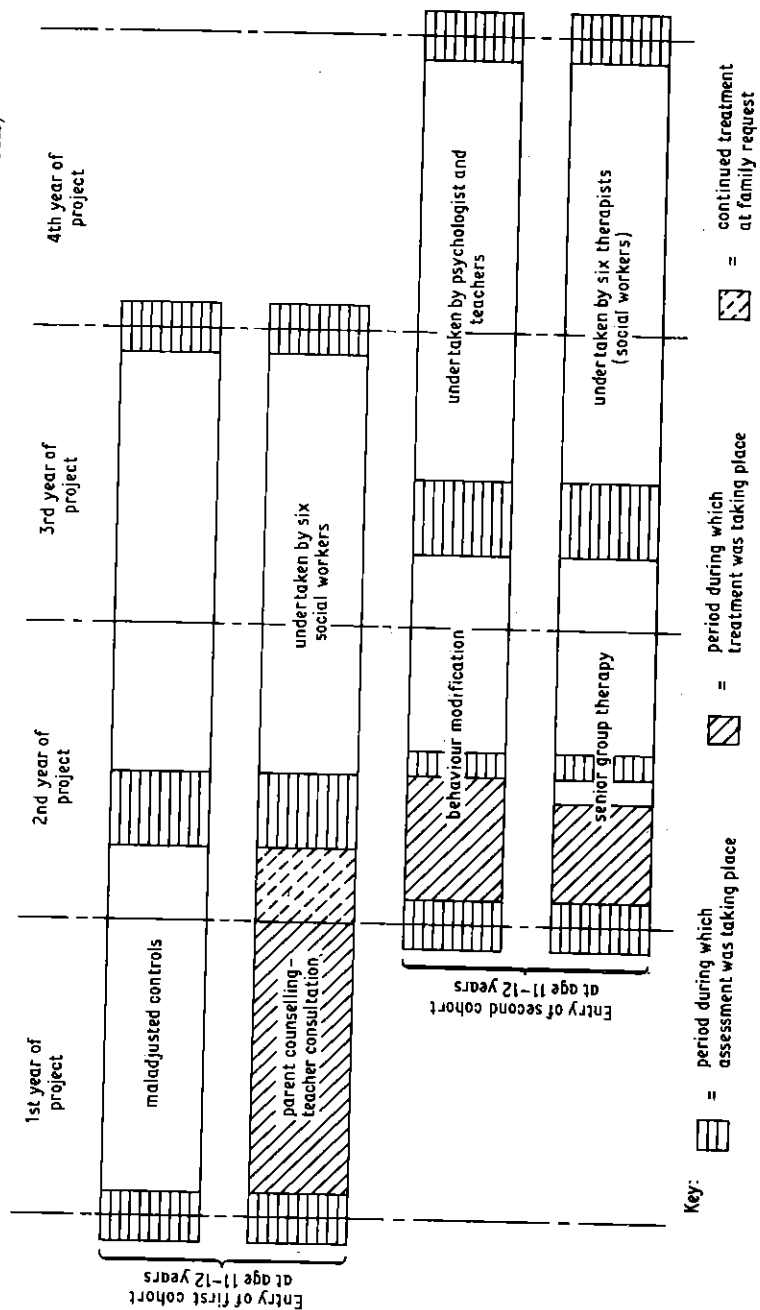


Figure 3(2) Flow chart of study: senior school programme (this programme took place in each of six senior schools)



### Summary of elements of treatment programme

In our project four forms of treatment were evaluated: parent counselling-teacher consultation in both junior and senior schools; group therapy with children in both junior and senior schools; nurture work with junior children; and behaviour modification with senior schoolchildren. For each treatment regime the senior members of the project team (who comprised mainly the authors of this book) organized training programmes for therapists and other personnel involved. Detailed training documents were also drawn up, where necessary. *Table 3(1)* shows the number of children selected for each treatment and control regime, and a full account of the types of treatment is provided in Chapters 5-8. A summary of the elements of the treatment programmes is given below.

#### BEHAVIOUR MODIFICATION

*Used:* in senior programme  
*Duration:* approximately two school terms  
*Personnel involved:* directed by a psychologist, implemented by teachers  
*Training:* introductory training manual followed by seminars  
*Programme:* defining and establishing goals; main technique of social reinforcement; individual behaviour prescriptions; continuous consultation with supervisor

#### NURTURE WORK

*Used:* in junior programme  
*Duration:* five school terms  
*Personnel involved:* carefully selected non-professional teacher-aides; teachers; mental health professionals  
*Training:* teacher-aides - some training but retention of natural style  
 teachers - explanatory seminars  
*Programme:* compensatory and enrichment activities; emphasis on healthy interaction experiences; some behavioural shaping; individually tailored help recommended for treated children, *plus* regular discussions and support from mental health professionals



## 50 Help Starts Here

### PARENT COUNSELLING-TEACHER CONSULTATION

- Used:* in both junior and senior programmes  
*Duration:* three school terms  
*Personnel involved:* school-based social workers; teachers; back-up team of senior social workers and psychiatrists  
*Training:* social workers having special training for their role in an educational setting  
*Programme:* school-based activities – consultation with teachers of the identified children; attempts to link home and school  
home-based activities – planned short-term case-work with families (averaging six sessions), plus back-up support for social workers from psychiatric team

### GROUP THERAPY/PLAYGROUPS

- Used:* in both junior and senior programmes  
*Duration:* one school term – ten sessions  
*Personnel involved:* social workers ignorant of information from any previous assessments of the children's problems  
*Training:* introductory training programme followed by continuous supervision by psychotherapist and child psychiatrist  
*Programme:* small groups – withdrawal from classes utilizing principles derived from Axline (1947a); Rogers (1952), and Ginott (1961)  
juniors – play and reflection of feelings  
seniors – more traditional therapy

### Evaluation of the effects of treatment

A series of follow-up assessments were undertaken at specified intervals, which included the end of the treatment (senior children only) and two main further assessments – the first being eighteen months and the final three years after the baseline assessment. We planned the intervals so that the time between assessments was brief enough to have been sensitive to any changes that occurred and yet, we hoped, long enough to allow changes to occur. The different initial levels of severity of psychological disorder between the regimes had also to be taken into account in the analysis because initial levels inevitably affect the results. As well as the objective assessments shown in *Tables 3(2) and 3(4)* some more subjective views of improve-

ment were gathered and used to provide descriptive comparisons, but were not used in the main statistical analysis.

Fuller accounts of how we dealt with theoretical and technical problems of classification, definition, measurement, and details of treatment are provided in subsequent chapters and have already been described, in part, elsewhere (Garside *et al.*, 1973; Harvey *et al.* 1977; Kolvin *et al.* 1975a; Nicol and Bell 1975; Hulbert, Wolstenholme, and Kolvin 1977). Full details of our method are provided in Appendix 2.

#### OUTCOME

We used two methods to compare the effects of the different regimes (including the no-treatment regime). The first, more simple method was to calculate the outcome (as defined by Sainsbury (1975) and described in Appendix 2) for each child and then find the number who showed good, moderate, and poor outcome for each regime. This was done for three global ratings: general, neurotic and anti-social behaviours. Significance between groups was tested using the well-established chi-squared test. It should be pointed out that outcome ratings tended to be more reliant on data collected at home than at school.

#### IMPROVEMENT

The second, more complex method was to compare regimes by using analysis of covariance. By this method average improvement scores for each regime were compared for every measure separately at each subsequent follow-up. The special feature of analysis of covariance is that differences between regimes in initial severity and other factors which may affect improvement are taken into account. In our research project we took initial score, general severity of maladjustment, non-verbal IQ, and an index of social functioning of the family into account; we used these 'covariates' because preliminary analysis suggested that these were the most important of a larger number studied. In addition, an index of family history of psychiatric illness was used as a covariate in the junior regimes. On some of these factors the children in the four regimes (both junior and senior) differed, even though we had allocated the children to regimes on a random basis. While these differences were probably due to chance rather than being systematic, we thought they should nevertheless be allowed for. This was the justification for the use of analysis of covariance (see Appendix 2).

In addition, we summed certain measures and carried out analysis of covariance on these. Measures were summated on the basis of the results of factor analyses. For example, with the junior children it was

## 52 Help Starts Here

found that there were five variables that measured neurotic disturbance, and these were added together (equally weighted) to provide a 'neurotic' score. Five other variables measured antisocial behaviour and likewise, these were added together to give an 'antisocial' score.

Analyses of covariance were carried out on the four regimes as a group (separately for junior and senior children) and we also compared each regime with every other, taking into account the number of regimes (see Appendix 2 for details).

### **Numbers of cases, losses, and missing data**

We selected more cases than were necessary for our purposes in order to leave a reasonable safety margin for coping with unexpected losses. Kolvin *et al.* (1977) and Macmillan *et al.* (1980) have provided detailed discussions of the number of cases selected by screens. In the junior school programme we allocated 270 children to our treatment and control regimes but, for various reasons, five of these were not used, and 265 children entered the project. In the senior school programme we allocated 322 children; thirteen were not used, and 309 children entered the project. At the end of the follow-up period, three years later, we remained in contact with at least 95 per cent of cases in both junior and senior school programmes. However, contact does not necessarily imply that complete information was obtained. As we had anticipated, more problems occurred with missing data than with missing cases. This was because extensive data were collected from multiple sources, which meant that information could have been available for a child on one particular assessment but not on another. This problem of missing data occurred across time (i.e., assessment points), between the main sources of information (namely the home and the school), and within these main sources. One way of depicting it was to compare key data across time: for instance, while *full* 'screen' data were, of course, available at base, they were not available for 4 per cent of the juniors at the mid-point and 8 per cent at the final follow-up; nor were they available for 5 per cent of the seniors at the mid-point and 14 per cent at the final follow-up.

Information from home was always more difficult to collect than school data. Another problem was encountered when children moved school, in which case we did not consider it reasonable to ask the new school to allow us to undertake sociometry for a single child. Yet another involved the self-completion instruments, where difficulties arose with children who persistently failed to attend school or were unpredictable in their attendance. Key home data were not

available for 11 per cent of the juniors at the mid-point and 15 per cent at the final follow-up, nor for 11 per cent of the seniors at the mid-point and 16 per cent at the final follow-up.

We were particularly concerned with missing cases, and those cases where data were missing. We studied these in terms of all school and home data available at the baseline and, perhaps surprisingly, no evidence was found that the missing cases differed to any extent from those who remained in the study. The more complex statistical analyses utilized data both across time and assessments and, therefore, reduced the pool of cases with complete data available (see Appendix 3). We checked our results in various ways, and found that this reduction made no difference to the conclusions of the research project. For instance, on certain key variables we studied improvement on all cases available at each follow-up and found the differences between the two methods (i.e. common data as compared to all data) were trivial. In the less sophisticated analyses cases were not debarred by minor absences of data and therefore the pool of cases was more complete (see Appendix 3).