Part Three
Treatment approaches

5 Behaviour modification approach

Summary

We review the literature describing previous work on modification of behaviour in the classroom and through this see that it may be necessary to encourage and train teachers to change their style of classroom management, so that they emphasize praise for good behaviour, rather than disapproval of bad behaviour. In some cases rewards such as sweets and toys may be necessary in addition to praise and attention from the teacher. Improved academic achievement is often, but not necessarily, correlated with improved classroom behaviour. Children's interpersonal relationships may also be improved by behaviour modification techniques. It must be remembered, though, that classroom conditions in this country are not equivalent to the model settings in which much of the successful American research work has taken place.

We go on to discuss related issues, such as whether changed behaviour is maintained after treatment and in different settings, and whether types of behaviour other than those specifically treated may also change.

We describe five main components in training teachers to modify children's behaviour in the classroom. These are: instructional methods; feedback; social reinforcement; token reinforcement; and modelling. In order to have the best effect on the behaviour of the children in their classes it is important that teachers maintain their new techniques.

We go on to describe the methods used in our own behaviour modification programme, which were designed to illuminate some of the issues mentioned above. The project took place over a twenty-

week period in ordinary secondary schools – the 'natural environment' as opposed to an artificial research setting. Relatively large numbers of children and teachers were involved: the seventy-two children were roughly representative of the range of problem children in the school population, and the thirty-nine teachers were also fairly representative. Unlike most previous workers in this area, we used multiple measures of behaviour change and carried out long-term follow-up assessments in order to determine the persistence of treatment effects. A further important feature of our study was an attempt to reduce teacher training and consultation time to an absolute minimum.

The treatment methods in our project made extensive use of techniques of social reinforcement – praise, attention, and approval from the teacher were encouraged rather than the use of material rewards, although these were given in a few cases. We were careful to define the types of problem behaviour being treated, to specify successive steps through which a child might be expected to progress, and to pinpoint ultimate objectives closely allied to observable behaviour. Teachers were also encouraged to identify causes and effects of misbehaviour. In the hope that the effects of treatment would be maintained, written reports and suggestions for each child were provided for the teachers at the end of the programme. Efforts to involve parents were unsuccessful.

We attempted to test the importance of the neurotic/conduct disorder classification in relation to children's changes in behaviour as evaluated by teachers. We looked at the child's social and personal adjustment, academic progress, and academic motivation. In a substantial percentage of cases improvement was noted in each of these four dimensions. While no firm conclusions could be drawn concerning the relevance of the diagnostic categories in this context, those comparisons in which significant differences or trends were found favoured girls and also neurotic children of either sex.

Assessment of teacher motivation suggested that children taught by well-motivated teachers had the best results. Direct observation in the classroom suggested that children's task-related behaviour improved as a result of treatment, although no change in teachers' behaviour was demonstrated by this method of observation.

Independent evaluations comprised: (a) psychiatric ratings of outcome; (b) covariance analysis of improvement; and (c) comparison of selected sub-groups of children. These comparisons, drawn at the three follow-up stages, showed that there were some changes in responses in areas that were not specific targets of treatment. Second, the effects of behaviour modification could be detected up to two-

and-a-half years after the end of treatment. Third, the effects of treatment were ultimately apparent in the home, as well as at school. Fourth, changes were apparent to the child, to his or her peers, teachers, and parents.

Introduction

Research into behaviour modification in the classroom has mushroomed in recent years, as it has in other settings with a wide range of child and adult problem behaviours. There are several reasons for this. The first, and probably the most cogent reason, is the effectiveness of behavioural techniques in classroom management in influencing different types of social and academic behaviour. Second, the techniques can easily be used by teachers, not only in one-to-one or small group situations, but also with the more typical classroom size of thirty or so children. Third, not only do behavioural techniques complement regular teaching activities, they can also be seen to have an affinity with basic educational principles – many of the teacher's concerns, such as specifying objectives for learning, arranging for these to be gradually and systematically approached, and harnessing pupils' motivation, are also those of the behaviour modifier.

With the early applications of behaviour modification, as with most new fields of therapeutic endeavour, successes tended perhaps to be exaggerated and magnified. While research continues apace, the early over-optimism has been tempered and the real difficulties that confront behavioural applications in natural settings, away from the rarefied atmosphere of the laboratory or well-funded research projects, are being confronted. Many questions are still to be answered. For example, how can change procedures be designed to provide the best opportunities for maintaining improvement and ensuring that it transfers across settings? Which types of behaviour are the most appropriate targets for intervention? What is the most effective way of teaching the application of behavioural techniques? Once taught, how can continued application be ensured? Can programmes be developed that will fit readily into regular educational settings where the availability of professional manpower and consultation is minimal? In such settings how does a school's organization and administration impinge on an intervention programme? These are some of the difficult issues that face current workers, and to which answers must be provided if the promise of behaviour modification is to be fulfilled.

A review of the literature

In this review we discuss previous results of behavioural intervention for inappropriate classroom behaviour, academic behaviour, and problems of social interaction, and then we examine some critical current issues in research. The review is not intended to be exhaustive and will highlight issues pertinent to the current study. Comprehensive reviews are available in Sherman and Bushell (1974); O'Leary and O'Leary (1976); Nietzel *et al.* (1977), and O'Leary (1978).

For a full description of the form taken by classroom behaviour modification programmes see Macmillan (1976), Meacham and Wiesen (1974), Macmillan and Kolvin (1977a), and the descriptive account of our own programme in this chapter. The essence of classroom behaviour modification lies in the precise and unambiguous definition of types of behaviour to be changed, the specification of objectives and how they are to be approached, and the restructuring of environmental influences, both antecedent and consequent, in order to bring about the desired modification of behaviour.

INTERVENTION FOR PROBLEMS OF INAPPROPRIATE CLASSROOM BEHAVIOUR

Probably the most widely used technique in classroom behaviour modification has involved changing the ways in which teachers make use of praise, disapproval, and the withholding of attention. This is not surprising, given that all teachers use some such elements of social reinforcement, even though the procedures are not always used adequately or appropriately.

Some researchers have characterized the classroom as a

> 'barren wasteland when one compares it with other normally reinforcing interactions. Most of the reinforcers in the system are highly formalized, such as grades and test scores. The control of social behaviour is achieved more as a function of threatened or applied aversive consequences than by positive social reinforcers.'
> (Patterson *et al.* 1969:15)

White's (1975) results suggested that, after the first and second grades, the use of teacher approval diminishes and is consistently exceeded by the use of disapproval. Similarly, a large survey of public school teachers in Florida showed that 77 per cent of their interactions with children were negative while only 23 per cent were positive (Madsen *et al.* 1970). A common explanation of these findings is that disapproval and reprimands, by tending to stop misbehaviour immediately (although temporarily) have become popular techniques

with the teacher. This seems to be an unfortunate trap because, in the long run, continued use of high disapproval rates may be quite ineffective in management and attention to undesirable behaviour may have quite the opposite effect to that intended (Madsen *et al.* 1968a; Thomas, Becker, and Armstrong 1968). For example, in the former study, with a teacher trying to control out-of-seat behaviour, an increase in 'sit-down' commands actually led to an increase in out-of-seat behaviour.

It seems, then, that a shift to more positive forms of control, with emphasis on increased use of approval, may require a major change in teacher behaviour. The technique of social reinforcement may appear deceptively simple, but the extent of change required, on the teacher's part, should not be minimized – teachers have their own established patterns of behaviour, their own reinforcement history, and these cannot be easily discarded and laid aside (Hall *et al.* 1968b, in the USA; Ward 1971, in the UK). However, many studies are now available that point to the benefits, for both teachers and children, that may accrue from the skilled and sensitive use of social reinforcement. Many of the early pioneering studies were conducted in pre-school settings, and illustrated the use of reinforcement (for example, praise, attention, or verbal approval) following desired or appropriate behaviour, or the delivery of a reprimand following undesirable behaviour. These studies suggested that social reinforcement techniques could be employed effectively in the reduction of crying (Hart *et al.* 1964) and in the development of more active motor skills (Johnston *et al.* 1966).

Early work of this nature provided a stimulus and a model for subequent investigators: similar techniques both continued to be applied in pre-school settings (e.g. Brown and Elliott 1965; Buell *et al.* 1968; Allen, Turner, and Everett 1970) and were extended to include applications with older children in ordinary primary and secondary schools. Madsen, Becker, and Thomas (1968b), for example, working in a public elementary school in the USA, chose two children with problem behaviour from two separate classrooms, one with twenty-nine children, the other with twenty. After a baseline phase, three experimental components – rules, ignoring disruptive behaviour, and praise for appropriate behaviour – that had been found effective in combination in an earlier study (Becker *et al.* 1967) were introduced sequentially. It was found that rules alone were ineffective, and the rules and ignoring phase was associated with increased disruption. When praise for appropriate behaviour was added to the two other elements, inappropriate behaviour fell from a 70 per cent baseline level to about 30 per cent. Reversal to baseline conditions produced a

corresponding increase in inappropriate behaviour, whereas re-instatement of all three experimental components restored satisfactory levels, thus demonstrating their effectiveness. Although the authors pointed to the success of combined praising and ignoring procedures and suggested that praise for appropriate behaviour is probably the key to effective classroom management, it must be stressed that their experimental design, by cumulatively combining procedures, rather than using them separately, made drawing clear-cut conclusions difficult. Nevertheless, this was a seminal study. Numerous subsequent projects have successfully exploited the systematic use of teacher attention to increase study, on-task behaviour, and desirable social behaviour (e.g. Hall, Lund, and Jackson 1968a; Hall *et al.* 1968b; Wasik *et al.* 1969; Broden *et al.* 1970; Breyer, Calchera, and Cann 1971; Peterson, Cox, and Bijou 1971). Other studies have focused on the reduction of disruptive conduct, such as noisy behaviour or talking in class, tantrums, and aggression (Thomas, Becker, and Armstrong 1968; Ward and Baker 1968; McAllister *et al* 1969; Hall *et al.* 1971; Lates, Egner, and McKenzie 1971).

As O'Leary and O'Leary (1976) have pointed out, the effectiveness of experimentally manipulated teacher attention has been demonstrated with a wide range of subject populations in a variety of contexts, ranging in age from pre-school (Schutte and Hopkins 1970) to twelfth grade (*c.* seventeen years) (McAllister *et al.* 1969), from classes of three children (Zimmerman and Zimmerman 1962) to classes of thirty-nine (Hall, Lund, and Jackson 1968a), and from normal classes (Madsen *et al.* 1968b) to classes for the retarded (Hall *et al.* 1971). Some studies focused on only one child in a class (e.g. Kirby and Shields 1972), while others included the behaviour of an entire group (Hall *et al.* 1968b).

We have already mentioned the possible undesirable effects of teacher disapproval: nevertheless, some of the studies cited above demonstrated its effectiveness when used in conjunction with praise and approval. Ignoring a child may not always be sufficient action to extinguish inappropriate behaviour, and stronger intervention may be necessary. McAllister *et al.* (1969) for example, achieved significant improvement in children's talking-out and turning-around behaviours with a combination of praise to the class as a whole for appropriate behaviour and disapproval to individuals every time they talked out of turn or turned around. Thus, selective use of disapproval may be beneficial and, indeed, such negative forms of intervention clearly cannot be ruled out when learning is being disrupted, or when a child may be hurt by another's aggressive behaviour. In this connection, Madsen *et al.* (1970) have suggested that it is the ratio between

positive and negative methods that is important. They advised a 4:1 positive to negative ratio as the optimum, although no data were presented to support this recommendation. Given that disapproval is necessary and inevitable, studies by O'Leary and Becker (1968) and O'Leary *et al.* (1970) suggested that soft reprimands, audible only to the offending child, are more effective in reducing disruptive behaviour than loud ones audible to the whole class. The latter may well give the miscreant an undesirable spot in the limelight, or create a 'ripple effect' (Kounin 1970) where surrounding children are affected by the scolding.

Most intervention involving social reinforcement has been in primary schools, and there is a relative dearth of studies with secondary schoolchildren. However, the behaviour of the secondary-school-age child is equally modifiable, as studies by McAllister *et al.* (1969), Atkinson, Davis, and Sanborn (1972), and Cormier and Wahler (1973) have demonstrated.

What is it that makes teacher attention so effective? Apart from its reinforcement or incentive value for children it may also have an informational function in cueing the child to behave in a particular way, especially when it is specific ('I'm pleased that you are asking more questions in class, Billy') rather than vague ('You've been a good boy today'). There is also the opportunity for observational learning by other children, so that the child who is praised acts as a 'model' of desired behaviour in the classroom (Broden *et al.* 1970). Much work remains to be done in identifying accurately the critical ingredients of social reinforcement. In addition to the suggestions above, other factors that need consideration are the proximity of the teacher, and the general effect of reducing disruption in the classroom. We have suggested earlier that the apparent simplicity of applying social reinforcement may be deceptive; it is equally clear that theoretical questions remain.

One fundamental drawback in the use of social reinforcement is that it may not be sufficiently powerful with some children. Teacher praise may, in fact, be aversive in some cases. In such situations the use of token procedures (rewarding with points or tokens exchangeable for a range of attractive reinforcers – sweets, access to preferred activities, choice of toys, etc.) in addition to the systematic use of teacher attention is likely to be more effective. It is worth stressing that it is desirable to make use of tokens in addition to and not instead of teacher praise and attention. Tokens, and the rewards to which they lead, are essentially props that facilitate behaviour change. The intention of such a programme, and the task facing the person administering it, is to bring the resulting behaviour change under the

control of natural variables in the child's environment, and appropriate use of teacher attention has a critical role here that must not be dismissed.

That token programmes are potentially more powerful than teacher attention alone was demonstrated in a study by O'Leary *et al.* (1969). Working in a second-grade class with seven problem children, aged about seven years, who were prone to wander around hitting other children and making a noise, they found that a combination of rules, altered classroom structure, and praise-and-ignore techniques were generally ineffective in reducing disruptive behaviour. Significant improvements were recorded with the introduction of a token programme. Many other reports are available that support the impact of token procedures in increasing task attention (Bushell, Wrobel, and Michaelis 1968; Broden *et al.* 1970; Sulzer *et al.* 1971; Ferritor *et al.* 1972) as well as reducing negative, disruptive, and deviant behaviour (O'Leary and Becker 1967; Carlson *et al.* 1968; Kuypers, Becker, and O'Leary 1968; Walker and Buckley 1968; Wolf *et al.* 1970). A number of classroom token programmes have taken the form of the 'good behaviour game', which employs groups rather than individuals (Barrish, Saunders, and Wolf 1969; Harris and Sherman 1973). The 'game' consists of dividing the class into groups that earn or lose points according to their behaviour. The group that earns most points (or loses least) wins the reinforcing event.

The use of token procedures in the ordinary classroom may seem alien and irrelevant to many (see Kuypers, Becker, and O'Leary 1968; O'Leary and Drabman 1971). However, in a useful recent review, McLaughlin (1975) has suggested that token systems can be readily adapted to regular classrooms. He pointed to the importance of such aspects as ease of implementation and management, low costs, compatibility with school and community attitudes, and pupil satisfaction in determining the viability of programmes.

It is possible that the behaviour of secondary-school-age children is more readily modified by token reinforcement than by the use of teacher praise and attention (Nolen, Kunzelmann, and Haring 1967; Meichenbaum, Bowers, and Ross 1968: Lovitt and Curtiss 1969; Williams, Long, and Yoakley 1972; Blanchard and Johnson 1973; Main and Munro 1975). Heaton *et al.* (1976), in a recent intervention with junior-high-school students, demonstrated the effectiveness of a token reinforcement programme (by comparing treated children with a control group in traditional schooling) in reducing misbehaviour and suspension from school, as well as in increasing academic achievement. Possibly the largest-scale application of token procedures was that undertaken in the USA by Rollins *et al.* (1974),

which incorporated secondary-school-age as well as elementary-school-age children. This was an ambitious study, involving over 700 subjects in sixteen experimental classes and fourteen control classes. Children in the experimental classes received tokens (exchangeable for sweets, toys, school supplies, and activities) for attentive classroom behaviour, and over the several months of the programme measures of disruptive and on-task behaviour showed their superiority over the controls in improvement of conduct. They also gained more than the controls on measures of IQ and academic achievement.

A further technique used to control inappropriate and disruptive behaviour is time-out, which involves placing the child in temporary social isolation following misbehaviour. Theoretical and ethical issues in its use were discussed by Macmillan, Forness, and Trumbull (1973) and Gast and Nelson (1977).

INTERVENTION FOR PROBLEMS OF ACADEMIC BEHAVIOUR

There are comparatively few studies demonstrating the effectiveness of teacher attention on academic achievement, which has more commonly been promoted through the introduction of token procedures. A rare exception is Stromer's study (1977) which successfully employed praise, with correctness feedback and modelling, to modify letter and number reversal difficulties in children from regular and special education classes.

Most classroom studies that focus on academic behaviour have employed some form of token reinforcement system. An early venture into the classroom by Wolf, Giles, and Hall (1968) showed the effects of token procedures in an after-school remedial programme for sixteen fifth and sixth graders (ten and eleven years of age) who were at least two years below reading norms and who were chronic low achievers. These children gained 1.5 years on the Stanford Achievement Test, while a matched comparison group gained 0.8 years in the same time. Similar impressive gains have been reported in remedial settings by Clark, Lachowicz, and Wolf (1968) and Kaufman and O'Leary (1972).

Hewett, Taylor, and Artuso (1969) rewarded an experimental classroom group of emotionally disturbed children with check marks (leading to tangible rewards like sweets, prizes, and extra time in arts and crafts) for being on time, following directions, and correctly completing assignments, and obtained significant gains in mathematics, but not in reading and spelling. In addition to improvements in basic reading and arithmetic achievement, gains in quality of handwriting, creative writing, and vocabulary have been reported (see O'Leary (1978) for review).

Token reinforcement systems can have as strong an effect on secondary-school-age children's academic behaviour as they had on their social behaviour (e.g. Glynn 1970; Chadwick and Day 1971; Kirby and Shields 1972). Sherman and Bushell (1974) suggested that no clear line can be drawn between investigations of classroom social and disruptive behaviour and those that concentrate on academic performance. However, much discussion has been directed at the question of whether it is best to modify social behaviour such as task attention, compliance, listening to instructions, etc., with the expectation that this will generate academic gains, or attempt to reinforce academic behaviour directly. This question is part of a wider debate within behaviour modification as to the appropriateness of behaviours selected as targets (Winett and Winkler 1972; Emery and Marholin 1977).

A number of workers have documented the extent to which academic achievement and social behaviour are correlated. Thus rating-scale behaviour such as attention, persistence to task, compliance with teacher demands, and ability to follow directions has been found to correlate highly with various achievement measures and teacher grades (Davidson and Greenberg 1967; Swift and Spivack 1968, 1969). Longitudinal studies have demonstrated the utility of teacher ratings of 'attention' and 'co-operation' in predicting later achievement (Meyers, Atwell, and Orpet 1968) and the extent to which aggressive and disruptive behaviour can forecast academic failure (Feldhusen, Thurston, and Benning 1970). Attention and attending behaviour are repeatedly highlighted as important to academic success (Lahaderne 1968; Bryan 1974). Cobb and his colleagues (Cobb 1970, 1972; Hops and Cobb 1973) argued that these are 'not academic behaviours per se, but rather, the first components in a chain of correct academic responding' (Hops and Cobb 1973:196). They view them as 'survival skills', necessary, though not sufficient, for successful academic functioning. If these survival skills can be improved and strengthened, academic gains will follow. This hypothesis has been supported for reading achievement (Cobb and Hops 1973; Hops and Cobb 1973) and maths (Walker and Hops 1977).

Evidence also exists, however, to show that increasing attentive or reducing disruptive behaviour does not necessarily have an effect on academic performance (Ferritor *et al.* 1972; Harris and Sherman 1974; Marholin *et al.* 1975). If one wishes to modify academic behaviour it may be best to reinforce academic behaviour directly in the form of gains on tests or achievement measures. This has been found to not only improve academic behaviour but also to increase attentive and decrease disruptive conduct (Ayllon, Layman, and Burke 1972;

Winett and Roach 1973; Ayllon and Roberts 1974). While the evidence here is conflicting, it is clear that the selection of target behaviours for modification must be carried out with care and with consideration of their potential value, if improved, for the child, as well as for the teacher.

INTERVENTION FOR PROBLEMS OF SOCIAL INTERACTION

In addition to coping with academic requirements and demands for behavioural control in their school life, children also have to acquire the skills that make for effective interpersonal relationships in their peer group. Poor peer relations predict subsequent disturbance (Combs and Slaby 1977; Macmillan *et al.* 1978) and it is therefore important to cope with these difficulties at an early stage.

Dissatisfaction has been expressed with methods in this area, which rely heavily on adult-administered techniques. Strain, Cooke, and Apolloni (1976) suggested that the methods require continuous teacher presence if the effects are not to disappear, and, additionally, that effects are liable to be beneficial only inconsistently as children move around a variety of classroom environments. Adults intervening to provide reinforcement for interaction may well interfere with and disrupt the interaction (O'Connor 1972) and considerable skill may be called for in providing sensitive reinforcement (Roedell, Slaby, and Robinson 1977). Some approaches depend on the occurrence of desired behaviour, or approximations to it, which can then be reinforced and gradually shaped. If relied on exclusively, these methods may be excessively time-consuming, or with extremes of isolated or aggressive behaviours, unfeasible. They are also limited where new or complex social skills need to be taught, such as the skills involved in initiating interaction or solving social conflict. Finally, adult-dependent methods disregard the critical contribution of peers in social interaction and the central role of the reinforcement that they provide.

A number of studies are available that overcome some of these difficulties. Modelling and shaping methods have been used successfully (O'Connor 1972; Evers and Schwarz 1973) but their results have not been consistently replicated (Gottman 1977).

Gottman, Gonso, and Schuler (1976) elaborated the modelling method by adding a social skills training package. This programme sought to teach skills that were found to play an important part in the discrimination between popular and unpopular children and that contributed to effective social functioning. Nine weeks after the intervention significant changes in sociometric position were reported for the subjects.

Allen *et al.* (1975) undertook behaviour modification in groups of sociometric isolates, meeting on a weekly basis about thirteen to seventeen times for sessions lasting fifty minutes each. The programme proceeded in three phases: (a) simple interactive games; (b) games emphasizing poise and flexibility within a social group; and (c) social play outside the group, in the playground. Children were reinforced with tokens for appropriate behaviour. Sociometric measures at the end of term and at follow-up five months later showed significant gains for the twenty-three treatment children, while untreated controls showed only minimal changes.

The value of taking the role of peers into account and stressing the reciprocal nature of interaction is apparent in the work of Walker and Hops (1973). Working with isolated children, they showed that interaction could be increased by rewarding (with points exchangeable for back-up reinforcers) both the withdrawn child and his or her classroom peers.

There are a number of procedures, not central to the concerns of the present study, that are receiving increasing research attention. These involve: (a) the use of group contingencies (Litow and Pumroy 1975); (b) the use of children as agents of change in the classroom, to modify both peer behaviour (e.g. Drabman 1973; Solomon and Wahler 1973) and teacher behaviour (e.g. Sherman and Cormier 1974); and (c) efforts to teach children methods of self-control, self-evaluation, and self-reinforcement (see Nietzel *et al.* 1977 for review).

The foregoing review demonstrated the potential impact of a variety of behavioural procedures in altering children's behaviour in the classroom. While the capacity of treatment to change behaviour is not in doubt, there are a number of issues in the literature on behaviour modification that deserve further attention and these are discussed below.

GENERALIZATION

The question of generalization is a major one for all applications of behaviour modification. It involves three basic issues. First, can changed behaviour be maintained after treatment ends? Second, can it transfer to settings other than the treatment setting? Third, can change be observed in behaviours that were not the specific targets of treatment?

Although a number of studies have demonstrated persistence and transfer of behavioural changes these are in the minority and reversal to pre-intervention levels has been the more typical pattern (Kazdin and Bootzin 1972; Marholin, Siegel, and Phillips 1976; Stokes and Baer 1977). Paradoxical as it may sound, generalization failures are

quite consistent with operant theory. If behaviour is a function of its consequences, then it will adjust to the influences prevailing in a given situation. If supportive reinforcement is withdrawn (as with termination of treatment or transfer to an unprogrammed setting) then the behaviour in question will probably decrease in strength.

MAINTENANCE

It is clear that there is a dearth of studies of maintenance in classroom work. Kauffmann, Nussen, and McGee (1977) surveyed 152 separate experiments conducted in classroom or educational settings between 1968 and 1974, and found that 72 per cent reported no follow-up data whatsoever. In the other 28 per cent, follow-up data were based on verbal reports (28 per cent), behavioural observations (54 per cent), or standardized testings (18 per cent). For studies relying upon verbal reports the average follow-up duration was 107 days, but only eighteen days for studies in which behavioural observations were employed. Thirty-six of the forty-two experiments in which follow-up data were reported showed effective maintenance. Less encouragingly, though, follow-up data were included in a greater percentage of studies published before 1971 than in those published subsequently. The paucity of follow-up studies is not restricted to classroom projects in behaviour modification but is prevalent to an equal extent in other areas, too (Cochrane and Sobell 1976; Keeley, Shemberg, and Carbonnel 1976).

The importance of securing control group data when maintenance is assessed was shown in an important study by Kent and O'Leary (1976). A standardized twenty-hour treatment programme involving the child, his or her parents, and teachers was evaluated, emphasizing social reinforcement from the teacher and special home-based rewards and privileges contingent upon school performance. Relative to no-treatment controls, the treated children had improved significantly on observational and rating measures by the end of treatment, but at the nine-month follow-up gains by the control group wiped out these differences. However, treated children showed better achievement scores and grades at follow-up than did the controls, though this was not maintained to the end of the study. Despite this long-term effect, it has been found in many classroom programmes that, once reinforcing contingencies are withdrawn, behavioural gains tend to dissipate (Birnbrauer *et al.* 1965; Kuypers *et al.* 1968; O'Leary *et al.* 1969) and behaviour change may not persist even into times of the day when programmes are not in effect (Kuypers *et al.* 1968; Meichenbaum, Bowers, and Ross 1968; Wolf, Giles, and Hall 1968).

TRANSFER

Few studies have been designed to find out whether a change in behaviour in one setting results in a similar change in other settings where a programme is not in operation. In educational environments it is important to know if programmes in one classroom have implications for children's behaviour in others, if gains in 'special' class or school settings transfer to regular settings, and if improved behaviour in school transfers to the home. As with maintenance, training effects have been found to tend not to transfer across settings (e.g. Kuypers *et al.* 1968; O'Leary *et al.* 1969; Broden *et al.* 1970). These discouraging findings are hardly surprising in view of the situation specificity of behaviour (Mischel 1968). Where problem behaviour exists in two settings, intervention may be necessary in both. Thus, Wahler (1969), working with two boys showing deviant behaviour in home and school, found that modifying behaviour in the home did not result in corresponding changes in school. Change did not occur in the classroom until procedures of differential attention similar to those applied in the home were extended to the school. Even with less dissimilar settings transfer may not occur. Glavin, Quay, and Werry (1971), in the USA, employed a token system to reduce disruptive behaviour and increase task-related behaviour in a special classroom for elementary children. The improvements obtained in this setting were not observed on return to the regular classroom, and behavioural and academic gains were not maintained at the two- to three-year follow-up (Glavin 1974).

CHANGES IN BEHAVIOURS OTHER THAN SPECIFIC TARGETS OF TREATMENT

The optimist who looks for maintenance and transfer of treatment effects also looks for a spread of effects to responses other than the ones targeted during a programme. Behaviour modification has perhaps been excessively concerned with measuring only responses that have been the focus of treatment; the need for greater attention to multiple response evaluation has been expressed recently (McNamara 1975). As well as being concerned with positive transfer to other behaviours we should also, of course, be aware of the possibility of negative side effects. Prompted by the latter concern, Ward and Baker (1968) found neither positive nor negative effects in the Wechsler Intelligence Scale for Children (WISC), projective questionnaire, and Draw-a-Person scores following a programme involving changes in teacher attention. However, Twardosz and Sajwaj (1972) obtained gains in social interaction when reinforcing in-seat behaviour. The studies examined earlier in this chapter, which show achieve-

ment of IQ gains following reinforcement of attentional or task-related behaviour, have provided evidence on the wider effects that may ensue from treatment, but these results are not always obtained. In a rare attempt to measure children's attitude changes associated with participation in a classroom behavioural programme, Buys (1972) found that children perceived their teachers as liking them more, became more positive in their own attitude towards the teachers, but evaluated being good in class in a more negative manner.

PROGRAMMING FOR MAINTENANCE AND TRANSFER

The message here seems to be clear, and it is conveyed in the frequently quoted but less frequently implemented suggestion that 'generalization should be programmed rather than expected or lamented' (Baer, Wolf, and Risley 1968:97). The picture is not entirely bleak, however, and with the growing concern with these issues methods for promoting maintenance and transfer have multiplied. Some positive findings are emerging, and these can be considered in the light of the particular techniques employed.

The first of these techniques involves the development of types of behaviour that are likely to be maintained by the natural environment, as reflected in the suggestion by Baer and Wolf (1972) of a 'behavioural trap': if behaviour can be developed, even through artificial or contrived means, it may then be naturally supported and maintained by the consequences (for example, praise and attention) it elicits. Allen *et al.* (1964), for example, found that after increasing a withdrawn child's interactions, removal of attention did not result in relapse – the behaviour may have been maintained as a function of the 'trap' of social interaction.

Second, attempts can be made to substitute one programme for another. For instance, Walker, Hops, and Johnson (1975), returning children to a regular class setting from a special education, token economy classroom, compared a substitute programme having natural reinforcers (praise, grades) with a control condition with no programme in effect. Children with the substitute programme maintained their behaviour better than did those without one and, when all programming was withdrawn, the former group's superiority was evident at the four-month follow-up. Substitution of one programme for another is not strictly a maintenance strategy – the programme still only showed that its implementation altered behaviour in the subsequent setting.

A third measure for promoting maintenance and transfer is to gradually remove or thin out the reinforcement being provided. After

controlling disruptive behaviour in an adjustment class with token reinforcement procedures, Drabman, Spitalnik, and O'Leary (1973) gradually faded out the teacher rating by which point-earning was determined. During a final phase in which no checks were made by the teacher – unfortunately a brief twelve days – appropriate levels of behaviour were maintained. Similarly, Turkewitz, O'Leary, and Ironsmith (1976) faded out back-up reinforcers and found maintained behavioural improvement after the reinforcers were completely withdrawn, albeit for the relatively short period of five days. Greenwood, Hops, and Walker (1977) obtained substantial maintenance over a nine-week period following cessation of formal intervention procedures by phasing out the specific behavioural programme materials, the classroom rules, and a class bar graph indicating daily progress. Phasing out may, to some extent, involve periods of intermittent or irregular reinforcement, a technique that is traditionally recommended where maintenance is sought. It may also be implicit in the fourth strategy of delaying delivery of reinforcement. Greenwood *et al.* (1974) gradually increased the number of sessions before reinforcement could be earned, and maintained appropriate behaviour at the three-week follow-up.

A fifth method for seeking maintenance of treatment effects, and one that is gaining in popularity, is the use of self-reinforcement. Theoretically, if a child can become relatively independent of external reinforcement and can control his or her own behaviour, gains may not disappear once programmes come to an end. However, when attempts are made to replace external reinforcement by self-reinforcement, excessive leniency in self-reward may result (Santogrossi *et al.* 1973). This is not always the case, however. In the study already cited, by Drabman, Spitalnik, and O'Leary (1973), children had complete control over reinforcement administration in the successful twelve-day phase when teacher management was eliminated. Bolstad and Johnson (1972) showed that self-administered points were superior to teacher-administered points in controlling disruptive behaviour, but the maintenance period was of only seven days' duration.

Finally, the involvement of parents seems essential if it is intended that effects of classroom-based programmes transfer into the homes. The extent to which parents control reinforcers for their children (for example, access to toys, play activities, TV, pocket money, etc.) underlines the potential impact of their involvement for generality of effect. Many studies have successfully involved parents to buttress school-based intervention (see Atkeson and Forehand (1978) for review), although they have not directly addressed maintenance/

transfer issues. These studies have commonly involved conveying information to parents concerning their child's behaviour in school, and then making available contingent rewards in the home. A great deal has been written about parental involvement, and the subject has been reviewed by Johnson and Katz (1973), O'Dell (1974), and Graziano (1977); its potential for the prevention of later problems, as well as for the support and enhancement of intervention conducted elsewhere, is enormous. As with other parent programmes (Chilman 1973), problems in enlistment, maintaining interest, and modifying attitudes and behaviour remain.

TEACHER TRAINING

We know that a number of techniques exist that may be successfully employed by teachers to modify children's behaviour in the classroom: the important question centres on how teachers may be most effectively trained to apply these techniques. Methods of training teachers have been reviewed recently by a number of workers (e.g. Sherman and Bushell 1974; Copeland and Hall 1976; Kazdin and Moyer 1976). While many training approaches have adopted multifaceted procedures, a number of components that have been isolated in certain studies may be identified. Kazdin and Moyer (1976), for example, detailed five dimensions: (a) instructional methods; (b) feedback; (c) social reinforcement; (d) token reinforcement; and (e) modelling.

Instructional methods

Instructional methods, which include lectures, discussions, and course work – all generally didactic in format – are perhaps the most widely employed approach, and their popularity is reflected in the large numbers of instructional manuals now available for teachers. However, there appears to be no reliable evidence that an instructional element leads to application of behavioural techniques rather than merely permitting acquisition of knowledge of such principles and methods (Bowles and Nelson 1976).

A number of training programmes have been reported that rely heavily, although some not exclusively, on an instructional format. The Consulting Teacher Model, for example, emphasizes behavioural principles, measurement procedures, and the systematic use of reinforcement techniques (McKenzie *et al.* 1970).

A similar programme is the Responsive Teaching Model for graduates that usually takes the form of ten, weekly, three-hour sessions (Hall and Copeland 1972; Hall, Copeland, and Clark 1976). Graduate students lead discussion groups of about ten people and

content covers recording and measurement procedures, experimental designs, the examination of learning principles and research studies, and supplementary lectures and films. Students are encouraged to apply what they have learnt by conducting experiments in their classroom and subsequently discussing them in the group sessions. It was in this way that Hall hoped students would incorporate Responsive Teaching procedures into their actual teaching situations. There is a clear awareness of the need to translate knowledge into practice in this model and, indeed, to provide the combination of practical, theoretical, and research skills that enable teachers to build on what they have learned during the course. The model's effectiveness is attested indirectly by the large number of studies that have flowed from it – Copeland and Hall (1976) reported that some sixty Responsive Teaching studies had been published, and some 2000 educators had been trained since the programme began. Both this and the McKenzie model utilize relatively traditional academic practices that are generally assumed to assist the development of rather generalized skills. In studies on neither, though, was there any analysis of specific changes in teacher behaviour as a result of their training experiences.

However, didactic procedures may be effective if supplemented with training in the situtation in which the techniques are to be applied. Thus, McKeown, Adams, and Forehand (1975) found that disruption in the classroom was lessened following participation in a laboratory group, but not when only written instructional material was provided. The laboratory group afforded opportunities for supervision and reinforcement of attempted applications, feedback as to the quality of performance, shaping of teachers' behaviour, and modelling of described performance by the experimenters.

Feedback

The use of feedback is implicit in some of the models already discussed, in the context, for example, of discussion of experimental applications of procedures and selective reinforcement of teachers' efforts. In this section it will be taken to refer to the provision of information about the adequacy of performance or knowledge of results.

Feedback in training exercises usually takes the form of verbal or written reports of behaviour. Cooper, Thompson, and Baer (1970), for example, used verbal comments to one teacher and written notes to a second, referring to the teacher's frequency of praise and failure to praise appropriate child responses. Feedback was given every ten minutes and successfully increased the percentage of time both teachers spent attending to the children's appropriate behaviour. The

teacher who received written notes postponed reading them until the end of the day and, interestingly, the procedure was less effective with her than with the teacher who received immediate verbal reports, possibly because of the delay in feedback. After feedback was eliminated, appropriate teacher behaviour began to decline, but data on maintenance were not collected. That the immediacy of feedback may be critical was also suggested by the failures reported in studies where graphical or verbal feedback was given at the end of the day (Rule 1972; Saudargas 1972; Cossairt, Hall, and Hopkins 1973) or every other day (Breyer and Allen 1975).

Social reinforcement

While classroom behaviour modification literature is replete with successful examples of the use of social reinforcement techniques with children, there seems to have been a remarkable reluctance to consider the experimenter's or consultant's relationship with the teacher in terms of the principle of reinforcement theory. These principles operate at *all* levels in the consultant-teacher-child interaction.

Few studies are available in which conscious use has been made of social reinforcement to enhance the effects of training. Cossairt, Hall, and Hopkins (1973) found that systematic changes in teacher behaviour were not produced by merely telling teachers after each class how often the students attended and the amount of praise given. However, praise for the teachers' performance markedly increased their use of praise in the classrooms, and student attending behaviour increased under this condition. McDonald (1973) contrasted the effect of praising teachers for their selection of certain behaviour change strategies with that of simply telling them which strategies to employ. Teachers who received praise tended to show more 'supportive' behaviours and use fewer reprimands or commands ('desist' behaviours) than those who received consultation without praise.

These studies suggested that explicit use of social reinforcement can be a potent influence on teacher behaviour. What is surprising is that so little use has been made of it.

Token reinforcement

As with social reinforcement, examples of the use of token reinforcement are somewhat limited. McNamara (1971) compared the effect of token reinforcement (points were exchangeable for cans of refreshment) with that of a response cost procedure (withdrawal of tokens) in altering teacher attention to appropriate and inappropriate child

behaviour. The contingent delivery or removal of points was effective in altering teacher behaviour.

Modelling

Modelling has not been evaluated extensively in teacher training. This is despite its potential impact in conveying explicit behavioural requirements, an impact at its most forceful (Ringer 1973) when modelling is conducted in the classroom where the teacher actually functions. In Ringer's study, an investigator (model) initially took major responsibility for the administration of verbal and token reinforcement, while the teacher observed. The teacher was gradually introduced and given progressively more responsibility for administering reinforcement, while the experimenter was phased out.

Kazdin and Moyer (1976) concluded that it may be necessary to employ a variety of procedures in teacher training rather than to rely on a single approach. Modelling or role-playing may ensure implementation more readily than didactic methods. Provision of reinforcement is important in maintaining changed teacher behaviour as well as in its initial modification. Brown, Montgomery, and Barclay (1969) and Cooper, Thomson, and Baer (1970) have shown that withdrawal of reinforcement for teachers leads to their behaviour reverting to pre-training levels; this, of course, has predictable consequences for the children's conduct.

RESEARCH ENVIRONMENTS VERSUS NATURAL SETTINGS

Despite the numerous successes reported in the predominantly American literature on behavioural modification, concern has been expressed that the almost exclusive research framework of the studies makes it difficult to reproduce their results in ordinary schools (e.g. O'Leary and Kent 1973; Rollins *et al.* 1974). Specifically, it has been suggested by Kent and O'Leary (1976) that features such as course credits, joint authorships in publications, opportunities to work for a higher degree, and frequent monitoring of progress by research personnel, constitute advantages for teachers that are not usually available in natural settings but that may contribute substantially to positive outcomes. In addition, ready availability of research funds gives more flexibility in the choice of reinforcers for children and, in turn, increases the opportunity of maximizing motivation.

An equally critical aspect of the published literature is the extent to which studies have been designed as demonstration models. In other words, the concern has been to show that behaviour modification, or a specific technique within the behavioural umbrella, actually works, and this has been mostly achieved not by models that closely replicate

'natural environment' conditions, but by ones that, to a considerable extent, depart from them. What are the 'natural environment' conditions? In the UK, at least, they are frequently characterized at the general level by inadequacy of professional manpower relative to the numbers of children in need of help (Kolvin *et al.* 1975a) and, from the behaviour modification perspective, by limited teacher training and consultation time, and limited funds. In addition, psychologists working in schools often have to work with teachers who are not hand-picked for the task and whose motivation is less than ideal. They also frequently have to forego the luxury of independent observers who can provide the sort of data that is the cornerstone of the behaviour modification approach.

EXPERIMENTAL DESIGN

As already pointed out, the effect of reinforcement procedures in classroom behaviour modification has been assessed predominantly by reversal or ABAB designs (alternating periods of treatment and no treatment). There are a number of important research questions that these designs cannot answer adequately, and control group designs are increasingly recognized as being important in clarifying issues such as the relative efficacy of different treatments and the magnitude and persistence of behaviour change attributable to particular treatments, and in providing controls where there is multiple treatment interference (Kazdin 1973, 1975; O'Leary and Kent 1973). The latter authors have argued cogently that behaviour modification research may have more social impact when designed in the form of large-scale projects with control groups, than even numerous, successful reversal studies with small numbers of subjects.

Methods used in our behaviour modification programme

IMPORTANT THEMES IN THE PROGRAMME

Our behaviour modification programme had a number of features that were designed to take into consideration and clarify certain critical issues. First, the intervention took place in regular classrooms in ordinary junior high or comprehensive schools, so the project was firmly based in the natural environment. The secondary school population has been relatively neglected in classroom behaviour modification research (Heaton *et al.* 1976; Macmillan and Kolvin 1977a). This reduced the need to make inferences as to the applicability of behaviour modification from work in special settings or with younger children.

Second, relatively large numbers of teachers (n = 39) and children (n = 72) were involved. With few exceptions (e.g. Rollins *et al.* 1974) previous studies have dealt with limited numbers of both teachers and children (two or three teachers dealing with up to a dozen children is fairly typical) with attendant doubts as to general applicability. The children in our study exhibited a range of problem behaviour and were not selected purely for disruptive conduct: they could be described as roughly representative of the range of problem children found in the school population. Similarly the teachers, although certain selection criteria were applied (see p. 105), comprised a fairly representative sample.

Third, multiple measures of behaviour change were employed, supplementing the direct observations of behaviour that are the most frequently employed measures in classroom research. Very few workers (e.g. Buell *et al.* 1968; Chadwick and Day 1971; and Mulligan, Kaplan, and Reppucci 1973) have used multiple measures in this area. The issue of response generalization in behaviour modification research is an important one that has been neglected, possibly as a result of narrowness of focus in choice of target behaviours.

Fourth, an important part of the study design, in addition to the use of maladjusted controls and comparison with other treatment regimes, was the provision for follow-up. Recent studies have begun to pay more attention to the need to assess persistence of treatment effects (Patterson 1974; Allen *et al.* 1975; Kent and O'Leary 1976) reversing the tendency to ignore the problem completely or to conduct a follow-up after such a short period that the results are of very little clinical significance (Turkewitz, O'Leary, and Ironsmith 1976). While some behaviour modifiers would aver that what happens after their intervention 'is a function of chaotic or unfortunate programs of contingency . . . that are out of (their) control' (Willems 1974: 160), we would agree with Willems's rejoinder that this may be construed as an evasion of responsibility and that behavioural interventions might unwittingly disrupt good things or set bad things in motion that become clear only after long periods of time'.

The fifth significant feature of the study, and one that has considerable bearing on the ability to reproduce the method in ordinary educational/clinical psychological practice, was the attempt to reduce teacher training and consultation time to an absolute minimum. Some reports of the amount of professional time required to set up programmes are rather discouraging for the psychologist with a heavy caseload, who can devote only part of his or her time to intervention. Abidin (1971), for example, reported that thirty hours of

a psychologist's time is required to establish an individual behaviour modification programme with a teacher who has no previous experience of the approach, and 150 hours is needed to establish a token economy programme. Such demands on professional time seem quite prohibitive and unrealistic, and more typical of the traditional analytical therapies. Tomlinson (1972) conducted a more feasible exercise, reducing consultant time to 2.4 hours per child referral and 4.2 hours per token economy, by consulting with groups of teachers, concentrating consultation time during early stages of the programme, and limiting data collection. Intensive workshops offer another alternative in reducing training time, and these have been extensively employed (e.g. Madsen and Madsen 1973; Rollins *et al.* 1974) but are not always administratively or practically possible. Limiting professional time devoted to training and consultation is an important consideration when setting up a behaviour modification programme in ordinary schools, but should not put the quality of the project at risk. Training teachers in the principles of behaviour modification and ensuring that procedures are adequately implemented are not easy exercises that can be achieved by a simple series of instructions, but ones that require careful planning and monitoring.

THE SELECTION OF TEACHERS

In each of the six schools the Headteacher was asked to 'sound out' a group of about six to eight teachers whom he or she thought (a) might be favourably inclined towards involvement in the treatment programme and (b) had substantial teaching contact with the classes concerned. A discussion was then held with each of these groups in order to explain more fully the techniques to be employed and the role and extent of commitment anticipated for the teachers. At the end of this discussion the teachers were each given a copy of a document introducing the concepts, principles, and techniques of behaviour modification (Macmillan and Kolvin 1977b) and were invited to attend a series of three, weekly, group sessions dealing with this treatment approach. While there was no obligation to attend these seminars, the group of thirty-nine teachers who started the sessions were not regarded as volunteers because their nomination by the Headteacher must inevitably have created a sense of commitment, although this no doubt varied in extent from person to person. This was an important issue in view of its implications for co-operation and motivation in implementation of the programme. The group consisted of twenty-four female and fifteen male teachers. All but eight had more than five years of teaching experience.

TEACHER TRAINING

As mentioned earlier, we felt that, in order for our method to be replicable, training and consultation time should be kept to a minimum. Training was in three phases and was directed by a consultant, in this case a psychologist. First, the teachers were expected to read and assimilate the basic document referred to above. Second, three, weekly, one-hour group meetings were held with each set of teachers, either during the lunch-break or after school hours, whichever was more suitable. In these sessions, there was discussion of the content of the document, some relevant studies from the literature were described, and some time was devoted to discussion of how one or two of the selected cases might be handled in a behavioural framework. There was a limited amount of role-playing, for example to demonstrate the use of combined praise-and-ignore techniques. In the third phase, consultations with individual teachers were begun, once the group sessions had been completed. In these meetings discussion was initially geared to the organization of individual behavioural programmes for the selected children taught by that teacher. While all available information was drawn upon in the preparation of a treatment strategy, the emphasis was always on behavioural functioning and how the child could be helped in the classroom, as opposed to other settings. Detailed personal prescriptions were given to the teacher for each target child. Once treatment strategies had been prepared and implementation begun, the content of these sessions changed to discussion of practical difficulties and progress in the application of techniques. These sessions were continued throughout the duration of the programme, lessening in frequency with time and as teachers were considered to be capably handling the treatment requirements. 'Training' was therefore considered to be an ongoing process; the consultation phase was extremely important, because it offered some hope of bridging the gap between acquisition of knowledge and actual implementation, which the document/seminar phases alone would have been unlikely to achieve. In addition, these individual sessions presented an opportunity for the consultant to reinforce the efforts of the teacher in applying the techniques, and to provide specific and focused feedback on the observational data.

The average number of consultations conducted per child in the course of the programme was nine (range four-thirteen). The total time taken up by the seminars and the consultations in the six schools was eighty hours, with a mean of 13.3 hours per school (range 11.6–15.7 hours). This refers to on-site time only and does not include

time spent travelling and preparing material for seminars and consultations, or written prescriptions.

PROGRAMME DESIGN

After a three-week baseline period, during which the seminars were held and baseline observations were collected, the intervention programme ran in each school for twenty weeks. The introduction of the programme to the schools was staggered: the six schools were taken in three pairs, so that, as the baseline phase ended for the first pair and they proceeded into treatment, the second pair began the baseline and so on.

TREATMENT METHODS

We decided at the outset to emphasize techniques of social reinforcement – relying upon the systematic and contingent use of teacher praise, attention, and approval – rather than techniques based upon the use of material or concrete rewards. As was anticipated, however, we found that children differed in their response to social reinforcement: for this reason, material rewards were, in fact, given in a limited number of cases.

There were several reasons for emphasizing social reinforcement. One of these has already been discussed: this is the difficulty that might be experienced in the transition phase at the end of a programme where concrete rewards were employed. Social reinforcers, on the other hand, occur more 'naturally' in the child's environment than do material rewards and may therefore be more conducive to generalization and maintenance of behaviour change obtained in the treatment phase.

Second, the natural-artificial dimension of proposed rewards is a very significant one in an educational context, with continuing debate about the merits of extrinsic rewards and whether or not the emphasis should instead be on 'intrinsic motivation', with children being rewarded for 'learning for its own sake'. While we personally favoured the use of material rewards (Macmillan and Kolvin 1977a) our role in this exercise was not one of crusade and conversion, but one that clearly had to accommodate the ideals and philosophies of the schools, represented by Headteachers and the various teachers with whom we worked. There were, in fact, wide differences in attitudes on this issue, some teachers openly favouring concrete rewards and welcoming their applications, others rejecting them out of hand. Our decision to minimize the use of concrete rewards was, finally, as much dictated by their artificiality in the regular, natural classroom setting as by objections raised by teachers.

Lastly, because in the majority of cases we were dealing with one or two problem children within a large group (usually around thirty), it was considered invidious to make material rewards openly available to these few children and not the others. Teachers were quick to point out that this would be unfair and possibly lead to counter-productive rivalry and jealousy. Consideration of cost precluded the extension of such schemes to the entire class. Although the social reinforcement techniques recommended were intended to function as a general classroom-wide management strategy, it was inevitable that slightly more attention was paid to target children than to others by the teachers concerned. However, this was generally considered to be less obtrusive than administration of material reinforcers to the select few. Some teachers were openly opposed to the notion that certain children should receive extra attention at the expense of others, despite the fact that, even in the absence of a systematic intervention strategy, they tended to give more attention to these children anyway.

DEFINING PROBLEMS AND ESTABLISHING GOALS

Our first task when confronted with a problem child was to attempt to describe and define the areas requiring change, in observable terms. Thus 'nuisance', 'troublesome', 'unruly', might be translated into 'keeps talking out of turn', 'gets out of his or her seat and wanders around', or 'pinches other children'. This sort of specific definition is essential if there is to be a precise focus for reinforcement, and it facilitates direct and relevant assessment of problem behaviour. Once the behaviour has been defined in this way, it is equally important to apply the same stringent criteria in defining goals and objectives for treatment. Targets such as 'well-adjusted' or 'self-actualizing' are too vague and nebulous.

Accordingly, in the early phases of consultation, much time was devoted to clarifying concepts of problem behaviour, and encouraging precise initial definitions, clear specification of successive steps through which a child might be expected to progress, and statements about ultimate objectives that were closely tied to observable behaviour. The notion of developing behaviour through successive steps was considered to be crucial to the success of a programme, both in modifying teachers' unwarranted expectations of change, and in conveying the need for gradual 'shaping' of behaviour towards the specified goal.

Teachers were also encouraged to view behaviour in an environmental context, for example, by identifying possible causes and effects. What happened before John hit Bill? Was he teased? What were the

consequences of hitting? Did Bill cry, or stop teasing? This kind of approach to the events or circumstances associated with behaviour is known as a 'functional analysis' – an attempt to identify what is functionally related to the occurrence of the behaviour (Yule 1977). It involves postulating a relationship between the problem behaviour and environmental events as a prelude to altering these conditions. Teachers were encouraged, in particular, to observe their own behaviour *vis à vis* the child in terms of its possible functional role, to vary it according to the treatment prescriptions, and to monitor its effects. There are obvious constraints, of course, on the extent to which a teacher can carry out such experimental variations in the classroom.

A major theme in the area of problem definition and setting of objectives was identifying the range of problems that could come within the ambit of the programme. Teachers' interpretations of what constitutes 'disruption' inevitably vary, and it is clearly impossible to establish absolute uniformity in this area. However, the basic types of behaviour detailed (see *Table 5(5)*) were discussed with all the teachers as a means of achieving some common basis of understanding. One of the basic objectives of the programme was to increase the amount of time that children spent 'on-task' or engaged in learning, and this aim could be pursued with most of the children to be treated. Teachers were encouraged to provide reinforcement not only for task attention and studying behaviour, but also for increased/improved output and achievement. In pursuing these goals, one is automatically engaged in reducing disruptive behaviour or activities that interfere with learning.

The screen measures (see Chapter 3) disclosed a variety of behavioural problems, creating the task of evolving ways in which these could be tackled within the framework of the classroom. For example, although the Rutter B2 scale is not a measure that focuses solely on classroom behaviour, nevertheless much of its content reflects conduct that could be modified in the classroom. For example, restlessness, difficulty in settling, disobedience, worrying about things, complaining of aches and pains, and unresponsiveness, are all behaviours on which reinforcement techniques can be brought to bear.

JEPI neuroticism is somewhat different in that the contents of the items of this scale are rather general, and not necessarily focused on the classroom (for example, 'Do you sometimes feel cheerful and at other times sad without any good reason?' 'Are you touchy about some things?' 'Do you worry about awful things that might happen?'). This generality proved less of a difficulty than was anticipated, in that

teachers readily found, in children with high neuroticism scores, worries and anxieties that they could work with in the classroom – this applied even in the case of seven children selected solely by the neuroticism criterion. These children showed behaviours such as shyness and excessive sensitivity, and tendencies to worry about performance, marks, homework, and so on. Again, these were responses within the 'reach' of a teacher prepared to use gradual shaping, modelling, and prompting techniques to encourage such children.

Sociometric screen data are concerned with success and failure in social interaction and these are probably determined to a greater extent by behaviour in the playground than in the classroom. Nevertheless, there are several methods relating to social interaction that can be explored in the classroom:

(i) shaping procedures for withdrawn children to encourage interaction where classroom activities allow it;
(ii) the creation of special groupings in, for example, project work, craft work, etc. so that desired behaviours can occur (drama activities in two schools created some excellent opportunities);
(iii) seating with compatible peers;
(iv) reinforcing peers for interaction, or for initiating contact;
(v) with rejected children, reinforcing their co-operative behaviour or friendly approaches, to encourage these to take the place of conduct that usually tends to elicit rejection in others, for example, boasting, teasing.

With all these strategies, of course, opportunities for intervention were to be seized by the teachers wherever they presented themselves – in or out of the classroom.

In addition, the reading measure among the second-level tests enabled poor reading and associated responses to be targets for reinforcement. There was thus an effort in the programme to confront the spectrum of presenting problems rather than to focus narrowly on one or two limited types of response.

THE APPLICATION OF SOCIAL REINFORCEMENT TECHNIQUES

In the implementation of social reinforcement procedures we were guided by the established findings of workers such as Becker *et al.* (1967), Hall *et al.* (1968), and Madsen *et al.* (1968), although their subjects were younger than those selected for our programme. Comparable work with secondary-school-age children was rare (McAllister *et al.* 1969).

The social reinforcement applied took the form of directed attention, of comments of praise or approval, smiles or a nod, or physical

contact such as a hand on the shoulder or a pat on the back, whichever suited the teacher's personal style and abilities. It is quite clear that some people experience marked difficulty in using social reinforcement techniques, whether in normal or therapeutic situations, and with some teachers considerable encouragement and support were required. It was also pointed out that what was reinforcing for any particular child could not be defined *a priori* – for example, physical contact from a teacher of the opposite sex might embarrass some children; similarly, the class bully might regard the teacher's approval as highly undesirable in view of his or her projected 'hard' image. Accordingly, teachers were encouraged to experiment with different forms of social reinforcement if no success was immediately apparent. It was also possible that the reinforcement might vary or wane in effectiveness over time with any one child. For example, the comment 'good girl/boy' might be effective initially, but would soon lose its effect if repeated excessively without variation – so the need for flexibility in this respect was also stressed.

When developing specific aspects of behaviour, reinforcement was to be made available following particular responses. For example, a highly distractible child might be given comments of approval when observed attending to task; a shy, inhibited child might be given comments of encouragement and a smile when he or she made an assertive response in a group situation. It was considered desirable that comments of approval should convey clearly to the child *why* he or she was receiving such positive attention (for example, 'I'm pleased that you kept on working on these sums, John') rather than being vague, non-specific statements (for example, 'You've been a good boy today'). The requirements for a programme such as this are described more fully elsewhere (Macmillan and Kolvin 1977b). To encourage teachers to make frequent use of positive social reinforcement they were asked to look actively for specific aspects of behaviour that could be rewarded: to try, in the case of badly behaved children, to 'catch the child being good' (Madsen *et al.* 1968a). Where possible, teachers were to attempt to reinforce behaviour that was incompatible with the conduct they were trying to eliminate. For example, a child given to talking to neighbours might be praised for reading a book quietly, or an aggressive child might be given approval for co-operative responses.

Directing comments of approval at a child implicitly conveys information or cues to other children as to how they ought to behave. The child who receives approval may thus function as a 'model' for the others. Teachers were encouraged to use this 'modelling' opportunity: for example, in situations where a target child was behaving

badly, an adjacent child who was behaving well could be praised.

Minor instances of disruptive or deviant behaviour were to be ignored, on the assumption that attention directed to such behaviour, even of a critical, disapproving variety, might tend to reinforce and unintentionally increase it (Madsen *et al.* 1968a). The use of disapproval was not ruled out, however, and was recommended where it was impossible to continue ignoring misbehaviour, such as when the learning situation was being disrupted, or when any child was put in physical danger. An important consideration here was the balance achieved between positive and negative controls, and a heavy emphasis on the former was recommended. Where disapproval was employed, soft, private reprimands were considered preferable to loud, public ones, which may have an adverse effect by creating the very situation that a wrongdoer desires – publicity and a spot in the limelight (O'Leary *et al.* 1970).

APPLYING A BEHAVIOURAL CONTRACT SYSTEM

In a handful of cases, where social reinforcement was insufficiently powerful to effect changes in disruptive behaviour, behavioural contracts giving access to a variety of concrete rewards or preferred activities were employed. The initial step was to discuss with the child why such a system was being developed. This involved examining the implications of disruptive behaviour for the child's own learning and for the functioning of the class in general. The contract was presented as a means of helping the child to plan his or her own behaviour consciously and of affording greater control over its consequences. The arrangement was regarded as a private one, between child, psychologist, and the particular teachers involved. The confidential nature of the exercise was essential if the administration of concrete rewards on an individual, rather than a whole-class basis, was to be viable.

A list of positive and negative types of behaviour was drawn up, these being tailored to the needs of the individual case. Points values were attached to each of these, with points being earned for positive behaviour, and deducted for negative behaviour. A card bearing these details was given to the child, so that there was no doubt about which types of behaviour would be considered positive, and which negative. Teachers awarded or deducted points on the basis of behaviour during lessons, informing the child of the details at the end of the lesson. Points totals were recorded on another card, which the child could keep.

The first points target was always easy to achieve, so that the child could make a good start, but subsequent targets became progress-

ively more difficult, so that back-up rewards became more distant in time and were gradually faded out. Points targets were always decided upon in discussion with the child.

Rewards were administered by either the psychologist or a teacher, and were made available as soon as conveniently possible after a target had been reached. They usually took the form of items such as pencils, rulers, felt-pen packs, notepads, or small toys such as plastic animals or soldiers. Activities were sometimes structured as rewards and, as far as the timetable allowed, children would be allowed extra time in favoured activities, such as reading magazines, cooking, or caring for laboratory animals.

PROVISIONS FOR MAINTENANCE OF TREATMENT EFFECTS

It is evident from reports in the literature that unless specific steps are taken to facilitate and encourage the maintenance of treatment effects, the changes obtained may not persist once the programme has ended (O'Leary and Drabman 1971; Levine and Fasnacht 1974; Walker, Hops, and Johnson 1975). This consideration was one that determined, to a great extent, how the reinforcement programme was developed and applied, because the most appropriate time for planning for maintenance is not when a programme is over but rather when it is being designed and constructed. The decision to emphasize social reinforcement was perhaps the most important element in this consideration.

In those few cases in which material rewards were given they were always accompanied by comments of approval and praise, and were gradually phased out towards the end of the programme. In fact, with all forms of reinforcement there was a reduction in frequency and intensity towards the end of the programme. This was guided by previous findings that behaviour may be more effectively maintained by intermittent rather than continuous reinforcement (Bijou and Baer 1978) and by a desire to lessen the contrast between treatment and post-treatment environmental conditions.

When the consultation programme ended, further written prescriptions for follow-up management were made available to the teachers, with written feedback on the children's progress in treatment. Because the end of the consultation programme coincided with the end of the school session, it was inevitable that, in the new session, most of the treated children would no longer have teaching contact with any of the core of teachers involved in the programme. Although this presented a serious difficulty as far as maintenance was concerned, it was unfortunately one that we were precluded from tackling effectively because of lack of resources.

OUR ATTEMPT TO INVOLVE PARENTS

Our attempt to involve the parents of treated children in group discussions of behavioural management was a venture designed to achieve a spread of treatment effects as well as their maintenance. Apart from encouraging the transfer of gains obtained in the school setting, parental involvement was also considered to be of potential value in providing continuity of reinforcement against the background of frequent change and inconsistency of management in the school. There were not enough team members to enable the parental exercise to take place at the same time as the school-based programme, so its feasibility was not explored until near the end of the consultation phase. Lack of resources also prevented a personal approach or home visit to the parents to explain the nature of their possible involvement and seek their co-operation. The approach was made by letter, with pre-paid reply cards but, although initial non-responders were sent a second card, only 12 per cent of the parents showed interest, so the project was reluctantly abandoned.

Assessment of the behaviour modification programme by teachers and project team

In addition to the more objective assessments that afford comparisons between the children in the various treatment regimes and the maladjusted controls, we obtained some other perspectives on change solely in relation to the behaviour modification regime. Thus, we (a) obtained teachers' evaluations of change in various aspects of child behaviour; (b) examined the relation of teacher motivation to outcome; and (c) took observational measures of child and teacher behaviour before and during treatment.

CLINICAL EVALUATIONS OF IMPROVEMENT IN BEHAVIOUR MODIFICATION CASES

Behaviour modification programmes in schools frequently involve a three-level or 'triadic' model as described by Tharp and Wetzel (1969). In this approach, adopted in the present study, behaviour modification techniques are applied by a 'mediator' or 'change agent', a teacher who has substantial day-to-day contact with the problem child. This mediator is advised by a consultant, who does not deal directly with the child. Evaluations of this model have concentrated on measures of change in observed behaviour in child and teacher but these can be usefully supplemented by assessments that are more subjective in nature – evaluations of change made by teachers

themselves. These are useful for a number of reasons. First, the efficacy of behaviour modification is critically dependent upon its delivery by such people and the evolution of a flexible and sensitive system of management must be guided by change agents' views. Second, the findings of observational measures alone, although statistically significant, may be discounted as trivial and irrelevant by teachers if their perceptions have not correspondingly changed. Their evaluations shed light on the clinical relevance of the intervention. Third, despite the conspicuous success of behaviour modification in educational settings, this approach does not always meet with general approval, and, in fact, often gives rise to controversy. Hence, feedback from teaching staff can yield invaluable accounts of emergent problems, thus influencing the manner in which techniques are subsequently introduced and implemented.

NEUROTIC/CONDUCT DISORDER AND SEX OF CHILD IN RELATION TO IMPROVEMENT

In the analysis of teachers' evaluations there is interest not only in the overall improvement, if any, of the children, but also in whether certain types of children improve more than others.

Despite the well-established findings from long-term follow-up studies that children with neurotic disorders respond better to treatment than those with conduct disorders (Shepherd, Oppenheim, and Mitchell 1966; Robins 1972; Kolvin *et al.* 1977), such diagnostic classifications have previously been considered of little relevance to outcome in behaviour modification. This neglect of previous findings stems largely from behaviour modifiers' views that diagnosis involves theoretical categories and explanations that are often far removed from actual behaviour and that explain behaviour in terms of sub-surface dynamics – what a person *is* rather than what he or she does (Stuart 1970). Even more important than the possible disparity between diagnosis and actual behaviour, for the behaviour modifier, is the perceived lack of relevance of diagnosis for a functional analysis of problem behaviour and for the manner in which a behavioural programme may be implemented.

Nevertheless, it may be worthwhile for research to address itself to the crucial question of whether different kinds of disorder respond differentially to behavioural treatment. The aim of the present section is to describe an attempt at testing the relevance of the neurotic/conduct disorder classification, and a child's sex, to outcome as evaluated by teachers. The relationship between diagnosis and independent measures will be dealt with in Chapters 9 and 10.

Method of testing improvement

We were concerned with four areas of potential change in child behaviour:

(i) social adjustment as expressed in relationships with peers and teachers;
(ii) personal adjustments as expressed in the child's own personal well-being, in so far as this could be judged independently of interpersonal functioning;
(iii) academic progress – achievement evinced by general classroom performance or test results;
(iv) academic motivation – task-directed effort and interest.

The final consultation session with the teachers was devoted to discussion of these four areas in relation to each child in the programme with whom the teachers were in contact. An attempt was made to get a consensus view of the child's progress on the scale shown below. It was not always possible to see all the teachers involved with one particular child at the same time, so some teachers were seen individually, but all the teachers involved in the programme were consulted. Different evaluations had to be reconciled by the consultant. The following rating scale was used:

0: no problem in this area prior to treatment and remains no problem;
1: marked improvement – no longer recognizable as a problem in this area;
2: moderate improvement – considerable improvement but not to the extent of 1;
3: somewhat improved;
4: no improvement;
5: worse.

It was not possible to rate maladjusted control children on these dimensions, so the present analysis was restricted to within-group changes. As not all of the children had problems in each of the four areas being evaluated the number of cases relevant to the comparison was noted. Children rated 0 were excluded from calculations.

Results

Table 5(1) shows the number and percentage of children who fell into each change grade for the four areas of functioning. For social adjustment it can be seen that by summing grades 1–3, fifty-six out of sixty-nine children (81 per cent) appeared to derive at least some benefit from the programme. For personal adjustment the figure was sixty out of seventy-one (85 per cent). The relative figure for academic progress was thirty-four out of sixty (57 per cent) and, for academic

motivation, forty-eight out of fifty-seven (84 per cent). These summed improvement rates nevertheless gave a rather rosy picture of outcome as the 'somewhat improved' category on each dimension contained the bulk of cases.

Table 5(1) *Distribution of teacher ratings of change on four dimensions (end of treatment)*

rating of change	*social adjustment (n = 69)*		*personal adjustment (n = 71)*		*academic progress (n = 60)*		*academic motivation (n = 57)*	
	n	%	n	%	n	%	n	%
marked improvement	7	10.1	8	11.3	0	0	2	3.5
moderate improvement	16	23.2	19	26.8	11	18.3	22	38.6
somewhat improved	33	47.8	33	46.5	23	38.3	24	42.1
no improvement	11	15.9	9	12.7	26	43.3	9	15.8
worse	2	2.9	2	2.8	0	0	0	0

Comparisons of boys and girls, and of children with conduct and neurotic disorders, focused on cases falling in categories 1 and 2 as against those in categories 3 and 4.

For academic motivation, girls improved significantly more than boys ($p < .05$) and there was a trend for children with neurotic disorders to do better than those with conduct disorders ($p < .10$). With academic progress there was a tendency for girls to do better than boys ($p < .10$). No significant results emerged from the comparisons for personal or social adjustment.

With mixed groupings, for example for girls with conduct disorders, the samples were too small to permit an adequate comparison. For academic motivation there was a trend for neurotic boys to do better than boys with conduct disorders ($p < .10$). Tendencies of the same order were evident for personal adjustment, with neurotic boys showing greater improvement than boys with conduct disorders and girls with conduct disorders tending to do better than similar boys.

When we examined the relationship between clinical ratings and teacher consultation data, we found a significant tendency for greater improvement on social adjustment and personal adjustment to be associated with a higher number of consultations (correlation significant at $p < .02$ for both). The corresponding correlations for academic progress and academic motivation did not reach significance.

DISCUSSION

It was encouraging that for each of the four dimensions a substantial percentage of cases was considered to have improved – over 80 per

cent in the cases of social and personal adjustment and academic motivation. The lower improvement figure for academic progress (57 per cent) and its disparity with academic motivation may be due to a number of reasons: (a) the scope for observed improvement in academic achievement may not have been adequate where frequent assessments were not employed; (b) motivation may genuinely have improved without a concomitant change in actual progress; (c) increased motivation may have been more apparent than real; or, as the more cynical may prefer, (d) academic progress may be less open to subjective interpretation than the other dimensions.

Improvement rates above 80 per cent compare favourably with traditional estimates of remission in the absence of treatment, which are usually estimated as two-thirds (Rachman 1973), and even more favourably with the recent analysis by Bergin and Lambert (1978) indicating a median spontaneous remission rate of 43 per cent (see Chapter 10). The fact that the present improvement figures were obtained in such a short period is also encouraging.

The finding that girls' academic motivation improved significantly more than that of the boys, and the trend for girls to show greater academic progress, possibly reflects the well-established tendency for girls to be quieter, more conforming, and superior in performance to boys, especially in reading and verbal skills (Maccoby 1966; Blom 1971). In addition, teachers' tendencies to hold lower expectations for boys (Brophy and Good 1974) may also have made some contribution to the finding.

With regard to the comparisons for neurotic and conduct disorders only trends have been obtained; these favoured neurotic children on academic motivation and personal adjustment. A more stringent assessment of the relevance of diagnostic category, allowing comparison with maladjusted controls, is discussed in Chapter 10. To summarize, the comparisons yielding significant differences or trends favoured girls, and neurotic children of either sex.

It must be acknowledged that the evaluations of change discussed here were subject to bias from a number of sources. First, it is possible that the teachers could have 'faked good' in their reports of change to the consultants. This could have stemmed from dissonance-reducing motives (a need to see a favourable rather than an unfavourable outcome from their investment of time and effort in the programme) or from a desire to respond in terms of what they perceived to be consultants' expectancies. On the other hand, though, some of the teachers may have had no such motivation, or even an opposite one. The teachers were not, strictly speaking, volunteers for the programme, having been recruited by their Headteachers. Although

they were all given the opportunity to drop out before the programme began in earnest, the Headteachers' 'recruitment' to some extent implied that their choice was not an entirely free one. Because motivation for the programme varied considerably, some teachers may indeed have been motivated to 'fake bad' in these evaluations.

A further source of bias was that the consultants themselves, having no less an investment in the programme and in trying to achieve a consensus with the teachers, may have influenced their views in a more favourable direction. The writers were aware of these possible distortions and sought to represent the teachers' perceptions as faithfully as possible. These qualifications apply to the overall judgements of change of behaviour and the sex comparisons, but not to those comparisons involving diagnosis, because diagnostic status was not known by the consultants at the time of evaluation.

Teacher motivation

THERAPIST VARIABLES IN BEHAVIOUR MODIFICATION RESEARCH

There has been a certain reluctance in behaviour modification research to deal with, or in some cases even to acknowledge, the potential contribution to treatment of factors relating to therapist characteristics, motivation, and role in therapy. Although such issues have been subject to considerable exploration in psychotherapy research (Parloff, Waskow, and Wolfe 1978), behaviour modifiers have avoided this area, possibly because of a reluctance to deal with the 'intangibles' of therapy (Wilson, Hannon, and Evans 1968) and a conviction, bolstered by demonstrations of the effectiveness of automated procedures (Lang, Melamed, and Hart 1970), that the techniques are effective in their own right. In recent years the importance of therapist variables has been increasingly recognized in the behaviour modification field and these aspects are attracting greater research interest than in the past (Devoge and Beck 1978).

TEACHER MOTIVATION IN CLASSROOM BEHAVIOUR MODIFICATION

Behavioural methods in the classroom invariably involve the training of teachers, whether intensively or superficially, to adopt the role of the 'therapist' who administers the requisite procedures. The question of teacher motivation for such an approach is of immediate concern in such ventures, especially where teacher motivation is not manipulable by direct rewards or credits, or where the management

intervention takes place in regular school settings rather than in research or 'laboratory' classes. We have discussed issues of teacher motivation elsewhere in some detail (Macmillan and Kolvin 1977b) and have considered how enthusiasm for a behavioural approach and willingness to change one's own behaviour and responses *vis-à-vis* pupils may be affected by a number of factors already elaborated in previous sections of this chapter. We can reiterate briefly. First, whether a teacher has volunteered or been recruited for a programme may crucially affect his or her commitment to it. Unwilling recruitment into such ventures may well generate resistance to the demands being made. Second, if a teacher regards teaching as his or her primary responsibility, with pastoral functions as a secondary concern, the active involvement required in behaviour modification may not be viewed as a relevant activity. Third, if the teacher's view of behaviour aetiology and influence does not correspond with the current environmental theory of behaviour modification, attempts at environmental reprogramming may not be seen as a useful therapeutic activity. Fourth, punitive and disciplinarian attitudes to child management may be quite incompatible with reinforcement principles. These are issues that may affect the teacher's motivation for behavioural intervention, that may limit his or her interest in and enthusiasm for such an approach, and, in extreme cases, may prejudice the entire intervention.

In a therapeutic model that calls for *active, consistent,* and *sustained* involvement from the person mediating the intervention procedures, it would seem that the person's characteristics that bear upon his or her involvement could be critically related to observed outcomes. We describe here a method of rating aspects of the motivation of teachers involved in the behaviour modification programme, and examine the relationships between these ratings and clinically rated changes resulting from treatment.

METHOD OF RATING ASPECTS OF TEACHER MOTIVATION

A number of dimensions that were considered to have a bearing on teacher motivation were constructed. For purposes of analysis, the three that were considered to have the greatest relevance to involvement in a behavioural programme were selected. These were: (a) teacher's attitude towards behaviour modification; (b) teacher's willingness to change own manner of responding to children; (c) teacher's implementation of techniques. The anchor points of these scales are shown in *Table 5(2)*. To obtain an indication of inter-rater reliability the teachers in one school, where two consultants were

Table 5(2) *Teacher motivation dimensions*

(A) *Teacher's attitude towards behaviour modification*
1 seems very interested to develop this part of his/her work – a major part of teaching role
2 good interest – makes some effort to work in this way
3 some interest – shows some reluctance or inability to develop this role
4 no interest – disagrees with principles

(B) *Willingness to change manner of responding to children*
1 very willing – enthusiastic about change
2 favourably disposed to idea of changing own behaviour
3 willing to change but needs much persuasion and continued support
4 unwilling – opposed to attempts to change behaviour

(C) *Implementation of techniques*
1 good evidence of implementation (from observer/feedback/concrete examples of discussion)
2 moderate evidence of implementation/occasional implementation
3 slight evidence of implementation
4 little or no evidence of implementation

involved consecutively, were rated and correlations computed. A similar rating exercise was conducted by two consultants engaged in a separate programme and the pairs of correlations were averaged by the customary method as described by Fisher (1941). The outcome, as shown in *Table 5(3)*, indicated moderate agreement. About four weeks before the end of treatment all the teachers were rated on the three dimensions by the consultants.

Each child was allocated a 'teacher motivation score' with reference to the score of the teacher who took his or her class, or an average score if the child had exposure to more than one teacher. These scores were then correlated with the clinical evaluations.

Table 5(3) *Teacher motivation dimensions: inter-rater reliability (n = 32)*

dimension	*inter-rater reliability*
(A) attitude towards behaviour modification	0.77
(B) willingness to change manner of responding to children	0.70
(C) implementation of techniques	0.69

RESULTS

Table 5(4) shows the correlations between clinical evaluation of positive change and teacher motivation scores. All correlations but one (willingness to change/academic progress) were significant.

Table 5(4) *Correlation between clinical evaluation of positive change and teacher motivation scores*

dimension	*social adjustment (n = 69)*	*personal adjustment (n = 71)*	*academic progress (n = 60)*	*academic motivation (n = 57)*
(A) attitude towards behaviour modification	0.36**	0.46***	0.30*	0.39**
(B) willingness to change manner of responding	0.33**	0.35**	0.22	0.28*
(C) implementation of techniques	0.42***	0.41***	0.33**	0.32*

Note: *p < .05; **p < .01; ***p < .001.

COMMENTS

While these results indicated that favourable motivation on the part of teachers was associated with better outcome, the possibilities of bias in the data must be considered. The consultant who was responsible for collating teachers' clinical evaluations also provided the motivation ratings. While the latter were completed before the clinical evaluation exercise, the extent to which *both* ratings reflect aspects of the consultant's own perceptions cannot be determined. Thus, factors such as the consultant's own perception of child improvement and liking of teachers may have played a part in determining the results.

Observational measures in the classroom

Obtaining direct measures of behaviour has been of high priority in the development of behaviour modification. This reflects the concern to focus on problem behaviour that is capable of measurement and, also, the need to provide a basis for evaluation of intervention. Continued monitoring of behaviour following the pre-treatment or baseline assessment affords ready detection of changes, and treatment strategies can be modified accordingly.

It has been stressed elsewhere that we sought to shape the behaviour modification programme in the light of perceived constraints in the 'natural environment' of the ordinary school setting, rather than trying to create an artificial 'research' or 'laboratory' model of limited general applicability. Our placing of an independent observer in the classrooms, however, was a concession towards research and measurement requirements, and we accepted that such

personnel are rarely available in ordinary school settings alongside teachers who are keen to apply behavioural methods. (If such teachers are concerned with measurement and evaluation they have to evolve recording procedures that are compatible with day-to-day teaching activities and that they can apply themselves). Given this departure, we acknowledged that our resources and the size of our sample could not allow us to match the concentrated coverage in observation achieved by other classroom researchers.

METHOD OF OBSERVATION

The observer, a female psychology graduate, gained experience in the use of the recording system before the beginning of the study, in pilot work conducted in a school not involved in the main research exercise. She was instructed not to interact with any of the children, and to try to merge into the background as much as possible. To allow the children to get used to her presence her first exposure to each classroom did not involve systematic collection of data. The children were told that she was a student who would be sitting in for a number of sessions to observe teaching methods.

The behavioural coding categories employed for children and teachers are shown in *Tables 5(5)* and *5(6)* respectively. The observer had a clipboard, rating sheets, and a stop watch. An illustration of the rating sheet is given in *Fig. 5(1)*, on p. 126. Child categories were pre-coded down the left-hand side of the page. The recording procedure required the observer to observe a child for ten seconds and then in the next five seconds record both the child's behaviour and the teacher's response to it (in terms of the categories for teacher behaviour). Thus each cell represented ten seconds' observation and conveyed information about the behaviour of both teacher and child. It was hoped that this manner of recording would illustrate the nature of typical interactions between them. Having recorded one child's behaviour in this way, the observer would then move on to the next child for the following ten seconds, then the next, and so on. Having covered all the target children, the cycle would begin again, and continue until the end of the lesson.

Observation was conducted for treated children only, in classes taken by teachers involved in the research programme. It was also restricted to the baseline and intervention phases. As the observer completed her contract at the end of the treatment period no follow-up data were collected. Had such measures been available, their interpretation would have been complicated by the transfer of the bulk of the children, by that time, to new teachers in the following session.

Table 5(5) *Direct observation: pupil categories*

symbol	*category*	*definition*
X	gross motor behaviour	getting out of seat; standing up, moving around out of chair; rocking in chair; disruptive movement without noise
N	disruptive noise with objects	tapping pencil or other objects; clapping, stamping feet; rattling or tearing paper; banging books on desk; (rate only if noise can be heard with eyes closed; do not include accidental noise, or noise made while performing X above)
A	disturbing others directly and aggression	grabbing others' work/materials; knocking neighbour's book off desk; destroying others' property; kicking; hitting; shoving; pinching; slapping; striking or poking with object; throwing object at another person; pulling hair; tripping
L	looking around, distraction (non-task)	turning head or head and body more than 90° to look at another person, or the back of the room; showing objects to another child; attending to another child; looking into space; (exclude when turning or attending to another child occurs when under teacher's instructions, or in academic context)
V	blurting out, commenting, and vocal noise	answering teacher without raising hand or without being called on; making comments or remarks when no question has been asked; calling teacher's name to get attention; crying; screaming; singing; whistling; laughing loudly; negative comments towards teacher; (must not be directed to another child but may be directed to teacher)
T	talking	carrying on conversations with other children when it is not permitted; (must be directed to a particular child or children)
D	unresponsive	no response to teacher when asked questions or to make a contribution to the lesson
R	task-relevant behaviour	on-task, e.g. writing answers, answering questions, listening, raising hand; other behaviours clearly permitted by the teacher in the carrying out of an academic task

Arranging for the observer to see all the children in the treatment regime, in six different schools, proved no easy matter. Timetable clashes, absenteeism, school outings, travelling time from one school to another, and so on, all posed problems, and led to the adoption of a flexible schedule for observations rather than a rigid and unchangeable one. It became apparent that the demands on the observer to cover all the target children in six schools were excessive. Accordingly, one school was omitted from the observational exercise.

Baseline data could not be gathered on one child because of absenteeism; fifty-five children were ultimately involved.

Two twelve-minute blocks of observation were available for analysis for each child in the baseline period, and six twelve-minute blocks in the intervention phase. Percentages were computed for the time 'spent' in each behavioural category as a function of total time observed. The category of particular concern for purposes of analysis was task-relevant behaviour. Having similarly computed percentages for teacher behaviour, ratios were calculated for approval/disapproval, in line with the thrust of the training programme. In view of the relative shortage of data for some teachers, for purposes of analysis these ratios were calculated for schools rather than for individuals.

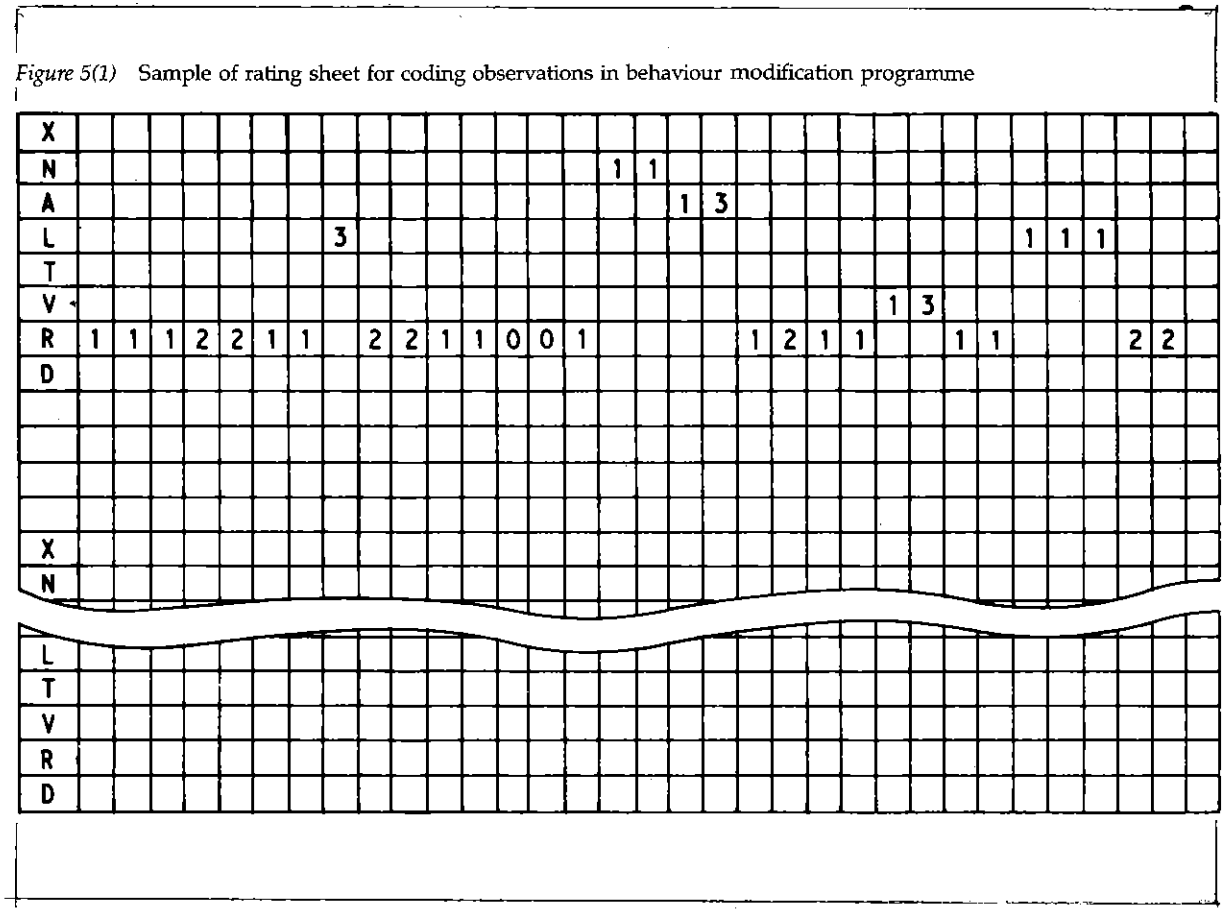
RELIABILITY OF THE OBSERVATION PROCEDURE

The reliability of the observation procedure was examined by our bringing in another observer who was using the same schedule in a separate project. The two observers simultaneously rated behaviour in a number of classrooms. Unfortunately, this could not be done before the beginning of the study, and took place during the

Table 5(6) *Direct observation: teacher response categories*

symbol	*category*	*definition and examples*
O	other response	calling on a child for an answer; probing for a response; non-critical instructions: one that does not imply disapproval
1	no response	the teacher makes no classifiable response to the behaviour observed
2	approval	(i) contact: positive physical contact such as patting, holding arm or hand (ii) praise: verbal comments indicating approval, commendation, or achievement, e.g. 'that's good', 'you're doing fine', 'you are studying well' (iii) facial attention: smiling at a child (iv) feedback: giving feedback for academic correctness
3	disapproval	(i) holding the child: forcibly holding the child; putting him or her outside; grabbing; hitting; slapping; shaking (ii) criticism: critical comments of high or low intensity; yelling; scolding; raising voice, e.g 'don't do that', 'stop talking', 'quiet!' (iii) threats: consequences mentioned by the teacher to be used at a later time: 'if . . . . then . . . .' comments (iv) facial attention: frowning, scowling, or grimacing at a child

Figure 5(1) Sample of rating sheet for coding observations in behaviour modification programme



intervention period. It involved observation of thirty-three children in eleven sessions, across the six schools, and a total of eight and a half hours' recording. Reliability proved highly satisfactory, ranging from 80 to 98 per cent agreement, using the formula:

$$\frac{\text{no. of agreements}}{\text{no. of agreements + no. of disagreements}} \times 100$$

RESULTS OF THE OBSERVATION PROCEDURE

The analysis of children's behaviour focused on changes in task-relevant behaviour. Rather than compare baselines figures with 'treatment' viewed in a global fashion, an attempt was made to detect changes within the treatment period itself. Thus, the treatment period was divided into three equal time-phases. Two observation blocks were available for each of these, and they were averaged to provide one value for task-relevant behaviour for each phase. The two baseline values were similarly averaged. Approval/disapproval ratios were also available for these separate phases.

A comparison of baseline task-relevant behaviour values and the three subsequent treatment values (phases 1, 2, and 3) was carried out by analysis of variance for repeated measures, for (a) all cases; (b) cases classified as 'conduct disorder'; (c) 'neurotic' cases, (d) boys; and (e) girls (see *Table 5(7)*). Where the analysis of variance proved significant the question of *which* means differed significantly was examined. (The statistical test used was the studentized range statistic q (Winer 1971).)

The above-mentioned values for each of the three phases of treatment and the baseline were compared, using the same statistical test as just described. The only pairs of means where significant differences occurred were as follows: when taking into account all

Table 5(7) *Observational data – task-relevant behaviour: significant differences between pairs of means*

cases analysed	*significant comparisons*	*significance level*
all cases	1 > B	$p < .05$
	3 > B	$p < .05$
neurotic cases	1 > B	$p < .01$
	2 > B	$p < .01$
	3 > B	$p < .01$
boys	3 > B	$p < .05$

Note: B = Baseline; 1 = treatment phase 1; 2 = treatment phase 2; 3 = treatment phase 3; the significance levels were derived by using the studentized range statistic q; > means 'better than' in this table.

available cases both the first and third phases of treatment differed from the baseline; for neurotic cases only, all three phases differed significantly from the baseline; and generally, in the case of boys only the third phase differed significantly from the baseline. There were no other significant differences.

Similar analyses were carried out for approval/disapproval values for the five schools. Again, comparing baseline and the three treatment phases, there were no significant differences.

DISCUSSION

The analysis of changes in task-related behaviour for the total group suggested encouraging gains as a function of treatment. In the absence of data on maladjusted controls, or a reversal analysis, this conclusion had to remain tentative. (Too many teachers had expressed concern about the successive application and withdrawal of treatment in a reversal design to make such an analysis possible, nor, indeed, had it been intended.) The trend of the means showed a pattern of rise, fall, then a rise in the final phase, with the first and last treatment phases, but not the second, being significantly higher than the baseline level. A possible explanation for this trend was that the teachers, or the consultant, or both, may have slackened off their efforts after the early phase of treatment, but, once it became evident that early gains were being lost, restored the previous level of commitment. Other explanations included short-term wash out of improvement subsequently compensated for by some longer-term learning process. The same pattern was not evident in the other significant analyses (for girls and neurotic cases) although, in both, the final phase was similarly superior to baseline. It is of interest that the trends reported in the discussion of clinical ratings (pp. 117–19), favouring girls and neurotic cases, were reflected more strongly here in observational data. The data suggested that the neurotic/conduct classification is of relevance to outcome in behavioural work, and it may be of value to workers to examine their respective behavioural referents in the classroom.

The gains in task-related behaviour were all the more encouraging, given that the baseline level was fairly high – 77.5 per cent. It was possible that this level was inflated because the teachers' seminars were being conducted during the baseline and they may have begun experimenting with procedures before the beginning of the treatment phase proper.

The most disappointing feature of the results was the failure to demonstrate change in teacher behaviour, via the approval/disap-

proval ratios, as a function of training. It may be that teachers' behaviour did not change in the forecasted direction, although the observer's anecdotal reports, and the consultant's impressions, suggested otherwise. On the other hand, it was possible that the observational procedure, perhaps by inadequately sampling teacher behaviour and not being sufficiently representative of it, was not sufficiently sensitive to detect changes.

This brings us to a consideration of other possible sources of error in the observational data reported here. Several reviews of the sources of difficulty in observation are now available (Lipinski and Nelson 1974; Kent and Foster 1977) and a number of critical issues have been identified. First, distortions may be produced by expectations on the part of the observer. Thus, O'Leary, Kent, and Kanowitz (1975) showed that data can be 'shaped up' by the observer when favourable comments are made by a consultant when shown data presenting the expected effect. Detailed and precise definitions of behavioural categories may help to reduce such expectation effects (Kent *et al.* 1974). In the present study the observer was fully aware, of course, that particular changes were desired in children's behaviour, so some degree of expectation bias may have been present. However, no discussions of data change took place between observer and consultant.

A second source of error is the reactivity of observational procedures – the possible effects on those observed of having a person present who is monitoring their behaviour. The nature of these effects is not clear: thus, for example, Surratt, Ulrich, and Hawkins (1969) suggested that the observer's presence leads to more on-task behaviour in the classroom, but Mercatoris and Craighead (1974) presented data showing no change in levels of appropriate student behaviour. We hope that the time allowed for habituation in the present study reduced unfavourable effects.

Third, 'observer drift' or 'instrument decay' may pose difficulties (Taplin and Reid 1973). These terms refer to deterioration in measurement attributable to changes in the observer, such as gradually changing definitions of behaviour, missing episodes of behaviour, adopting short cuts by collecting only parts of the data, etc. Random or covert checks on observer agreement help to control this. These were not possible in the present study, but definitions were regularly reviewed.

Finally, inter-rater reliability checks may produce inflated agreement scores when observers are aware that their data are being checked (Romanczyk *et al.* 1973). Checking agreement at unpredictable or unknown times may produce more representative data.

There is a growing awareness that direct observation is not necessarily a reliable, objective, non-reactive form of data gathering, immune from the problems of other kinds of measurement. Observers are fallible too. While sources of observer error have probably not been eliminated in the present exercise, they were taken into account and controlled, as far as resources allowed.

Independent evaluation

In this section we detail three different examinations of how children given behaviour modification treatment compared with maladjusted controls. First, outcome was based on clinical assessment of psychiatric status, yielding comparisons for base to midline, and base to final assessment. Second, we considered improvements on a wide range of behavioural and cognitive measures. These were assessed by covariance techniques which allowed for initial baseline differences between maladjusted control and behaviour modification children. Third, we focused on selected sub-groups within the treatment and control regimes, to make comparisons for which the total group method was not appropriate.

OUTCOME *(see Figs 5(2)–5(5))*

Considering first of all the global rating of severity, no difference was found between children in the two regimes in terms of good outcome for the base to midline comparison, but a striking difference in favour of behaviour modification was apparent for the base to final comparison ($p < .01$). The same picture emerged for antisocial behaviour, though the differences were less marked ($p < .05$). The controls were completely static on this measure. With neurotic behaviour, the behaviour modification children showed superior outcome for both comparisons (base to midline, $p < .05$; base to final, $p < .01$). These results reflect good maintenance effects for behaviour modification across time.

IMPROVEMENT – COVARIANCE ANALYSES

Covariance analyses of improvement for the behaviour modification and maladjusted control children will be considered for data relating to (a) the school and (b) the home. Only significant differences are reported here.

School-based measures

Baseline to end of treatment asessment. On measures relating to classroom behaviour significant differences were present for two Devereux

factors. The behaviour modification children showed greater creative initiative ($p < .01$) than did the maladjusted controls, scores that Spivack and Swift (1967) suggested were positively related to achievement, measuring 'the degree to which the child exhibits active personal involvement in and positive motivation to contribute to the classroom learning situation' (Spivack and Swift 1967:17). The treated children also showed an increased need for closeness to the teacher

Figure 5(2) Overall severity: seniors: per cent outcome (good and poor categories only)

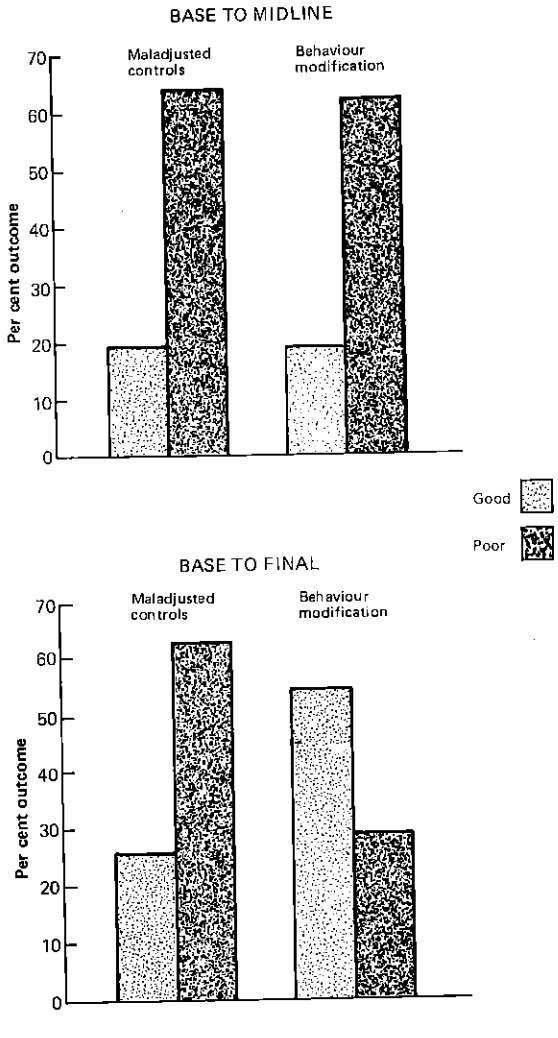


Figure 5(3) Antisocial behaviour: seniors: per cent outcome (good and poor categories only)

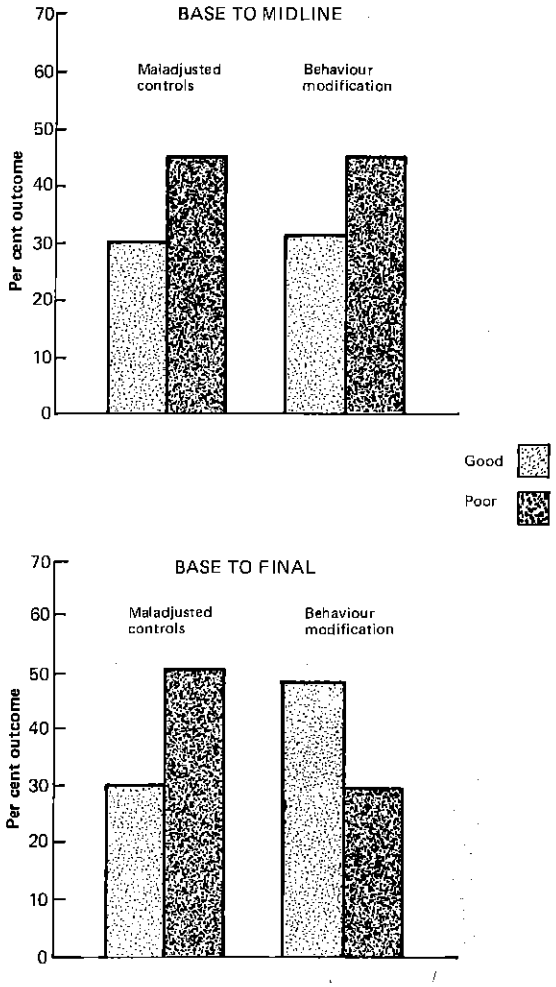


Figure 5(4) Neurotic behaviour: seniors: per cent outcome (good and poor categories only)

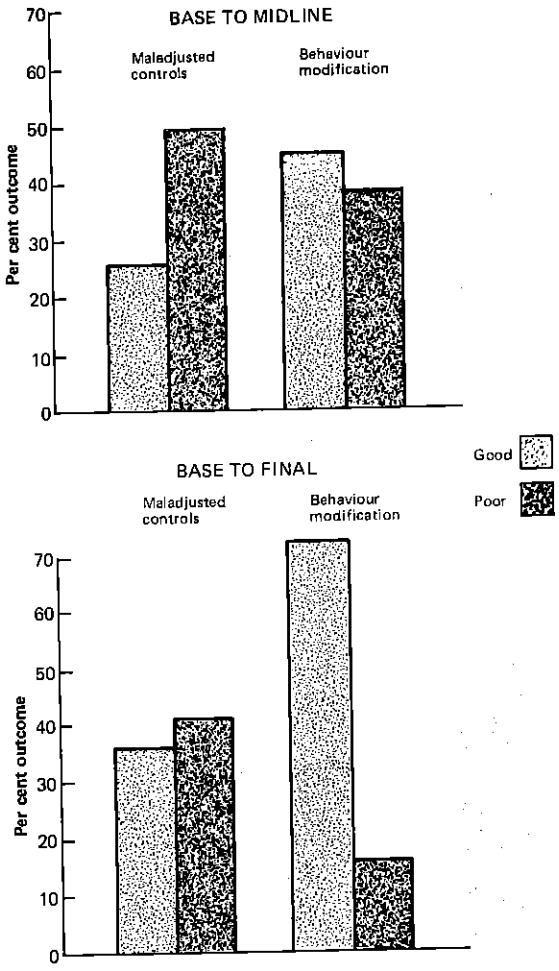
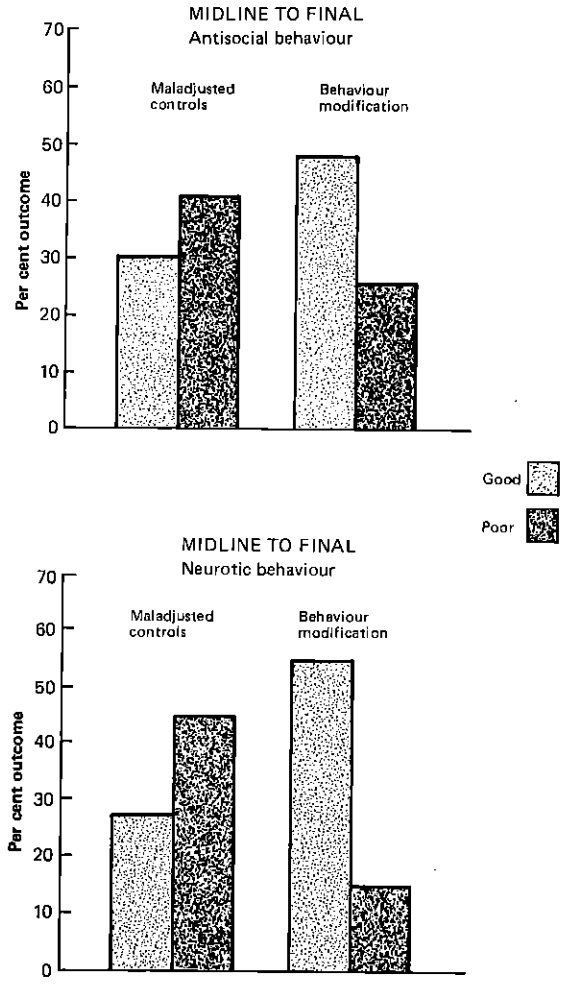


Figure 5(5) Seniors: per cent outcome (good and poor categories only)



($p < .01$). This factor reflects the extent to which children like to be close to, seek out, and offer to do things for the teacher – high scores suggest the teacher is positively valued by the child.

The behaviour modification children also became less isolated ($p < .05$) as indicated in sociometric measures, and improvement in attitude was also apparent, with the treated children showing a more positive attitude to school ($p < .05$) on the Barker Lunn scale (item 1) than did the maladjusted controls.

Baseline to midline assessment. None of the significant differences observed at the end of treatment was present in the midline comparisons. However, the behaviour modification children showed decreased neuroticism on the JEPI ($p < .01$), and also had significantly better scores than the maladjusted controls on Barker Lunn data, indicating decreased anxiety in the classroom ($p < .05$) and generally in relation to school ($p < .05$). Finally, changes on cognitive data favoured the treated children, with differences on verbal ($p < .01$), non-verbal ($p < .05$), and total ($p < .01$) ability scores.

Baseline to final assessment. Again, with this set of comparisons the pattern of differences changed. Behaviour rating scores showed significant improvement for the behaviour modification children on the Rutter B2 total score ($p < .05$), the neurotic sub-score ($p < .05$), and a decrease in JEPI introversion ($p < .05$).

Home-based measures

Baseline to end of treatment assessment. The only significant difference to emerge here was on the Rutter A behaviour rating total score, where the maladjusted controls did significantly better than the treated children ($p < .05$).

Baseline to midline assessment. Maladjusted control superiority was still evident here on the Rutter A total score ($p < .05$) and a further difference in favour of them was apparent on the Rutter A antisocial sub-score ($p < .05$).

Baseline to final asessment. The only significant comparison here was on antisocial behaviour (B), and this favoured the behaviour modification children ($p < .05$).

Aggregate measures (baseline to final assessment)

Comparisons on the global maladjustment score and the neurotic behaviour score were significantly in favour of the behaviour modification children ($p < .05$ and $p < .01$ respectively).

Improvement – sub-group analyses

Where specific kinds of intervention were developed for particular children, the total group comparison methods previously described may not be thought the most sensitive way of showing the effects of treatment. Given the multiple criterion method of selecting cases it was possible, for example, for a child to be selected solely on the basis of high teacher ratings, and to obtain scores indicating high popularity on sociometry. No intervention for peer difficulties would therefore be required, yet the total group comparison method included such cases, possibly obscuring changes in the core group with peer problems to which help was conveyed.

This comparison was therefore focused on children in the treatment regime who obtained extreme scores on the screen criteria – scores by which they would have been picked out from the total group as needing special help. They were compared with children selected from the maladjusted control regime in the same way. Scores were those above the cut-offs: Rutter, 9 and above; JEPI – N, 20 and above; Isolation, 0 or 1; Rejection, 12 or above. For teachers' ratings on the Rutter B2 scale, 't' tests showed the behaviour modification children to do significantly better than maladjusted controls between baseline and midline ($p < .05$) and between baseline and final follow-up ($p < .01$). For JEPI – N, behaviour modification was superior at the end of treatment ($p < .01$) and at the final follow-up ($p < .01$), but not at the midline. For the two sociometric indices, although changes at all comparison points were in a favourable direction, none reached statistical significance.

DISCUSSION

These comparisons at the three follow-up stages (end of treatment, midline, and final follow-up), while showing a shifting pattern of changes that makes an overall summary difficult, pointed to some interesting effects. First, there was evidence of changes in responses that were not the specific targets for treatment (for example, aspects of attitude to school, the personality dimension of neuroticism, and home behaviour). Second, there was evidence for the effects of treatment, not only at the midline follow-up but also, more impressively, two-and-a-half years after the end of treatment. Third, effects of treatment were apparent in settings other than those in which intervention was conducted. Thus the school-based programme produced some positive effects in the home, although not until the final follow-up. Fourth, in examination of sources of the measures – self, peers, teachers, and parents – there were no domains in which change was not reported at some stage. Fifth, it was only at

the final follow-up that there was improvement on aggregate measures of maladjustment.

The emphasis in the present study on using multiple measures of change and conducting long-term follow-up, sets it apart from most previous studies of behaviour modification in the classroom. The spread of treatment effects, and their apparent long-term effects, were a strong argument for this change in emphasis. It is pertinent to ask whether previous research has failed to identify important changes, either by not measuring them, or by not measuring them for long enough. The long-term effects were particularly encouraging in view of the fact that, at the points at which these data were gathered, the majority of the children were no longer in contact with the teachers involved in the programme, and provision for maintenance was minimal: new teachers had access only to written feedback on children's progress as a result of treatment, and brief written prescriptions for management were added to these. While these positive findings for generalization were encouraging, a price had to be paid, in that, having carried out the experiment in a natural setting, many questions about the specific factors controlling the process of generalization had to remain unanswered.

It is interesting that long-term changes tended to reflect 'sleeper' effects. In other words, effects appeared late, rather than at the end of treatment and being maintained over time. This indicated a need to re-examine the prevalent notion that changes in behaviour modification are very treatment-specific. In terms of teacher-report measures, this sleeper effect helps to rule out alternative explanations of change in terms of 'demand' factors (teachers being encouraged to report favourable change as a function of their involvement and investment in the programme, or in line with what the consultant might view as desirable.) Thus, while the end-of-treatment rating scales were filled in by teachers involved in the programme (yielding two significant results on the Devereux and on the Rutter B2 total for the sub-group analysis), the significant Rutter B2 results two-and-a-half years later came from teachers, the great majority of whom had had no involvement in the programme. Demand effects again seemed an unlikely explanation of parent ratings at the final follow-up, in that there was no convincing reason why they should have operated at this moment, not earlier. In general, the implication of demand effects is that change is being reported where none, in fact, actually occurred. However, if one takes the interactional view that children's problem behaviour is behaviour that comes into conflict with the environment, so that, in this context, teachers' and parents' perception of behaviour as deviant or problematic is a critical factor, then a treatment

programme that leads teachers and parents to perceive the child as improved has achieved some success.

It is intriguing that a favourable change in the parental reports of the children's behaviour at home did not emerge until the final follow-up, and that at the two earlier assessment points treated children were faring less well than the controls. How can this effect of poor progress at home, alongside beneficial changes at school, be explained? Other workers have found similar 'contrast' effects, with gains in one setting set against negative results in another (Meichenbaum, Bowers, and Ross 1968; Johnson, Bolstad, and Lobitz 1976). One possible explanation is that the children realized certain types of behaviour led to reinforcement in the school setting and they were accordingly motivated to continue this state of affairs; at home, reinforcement may have been absent and the children may have been disinclined to improve their behaviour, or, if frustrated by the lack of reinforcement, they may have increased their misbehaviour.

One of the most encouraging aspects of the positive results reported here was that they were achieved by low-key intervention, emphasizing natural forms of reinforcement, and applied within the constraints of regular secondary school settings (Macmillan and Kolvin 1977a). A major constraint concerned the 'reach' of the teachers' intervention. In the context of the present programme this had to be classroom-bound and tailored to be compatible with the teachers' primary task of teaching. To have had greater impact, more flexible intervention might have been necessary, allowing more individual approaches, and extending from the classroom to the playground, a significant arena in children's learning. Whether these developments could be achieved on the basis of resources similar to those of the present project is a question that will have to await further study.