

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

THE DEVELOPMENT OF SELF MONITORING OF BEHAVIOUR
BY DELINQUENT GIRLS IN RESIDENTIAL CARE

A Thesis presented in partial fulfillment of the requirements for the Degree of Master of Arts in Psychology
at Massey University

JOHN CHARLES WATSON

1981

ABSTRACT

The use of the methods of applied behaviour analysis to provide the basis of treatment programmes for delinquent youth has expanded over the past decade. These techniques appear to be adaptable to a wide variety of residential and community based settings. The token economy probably enjoys the widest use of all the programme alternatives. It has been found however, that when token programmes have been withdrawn, the maintenance of the behaviour change typically deteriorates. In an attempt to overcome this, self evaluation procedures can be an effective aid in maintaining behaviour after reinforcement contingencies have been withdrawn. This thesis reports on the introduction of behavioural self evaluation into a token programme with delinquent girls in a residential setting. In order to develop an accurate behavioural self evaluation and reporting procedure which established and maintained appropriate levels of behaviour, a matching procedure initially with contingent points for both accuracy of matching and behavioural performance was used. Girls rapidly learned to accurately assess their behaviour and maintain both assessments and behaviour at criterion levels. The matching procedure was withdrawn after criteria levels of 80 percent accuracy were maintained. Random checks that followed revealed accuracy and behavioural performance were maintained during periods of contingent reinforcement, and for a short period after reinforcement was withdrawn.

ACKNOWLEDGEMENTS

The author is extremely grateful to the staff members of the Margaret Street Girls' Home in Palmerston North, who gave their continued support, encouragement and criticisms of this research project.

To David Hutchinson, Barbara Sherson, Joyce McLaughlin, Betty Fraser, Ruth Grant, Betty Neilsen, Dawn Alexander, Jan Smith, Graeme Munford, Anne-Marie Van Der Linden, Maurice Field and Shelley Hamilton, thank you for all your work.

Special thanks also to the 28 girls who participated in the project, the standard of behaviour achieved was a credit to you all.

Thanks are also due to John Hancock, Director of Social Welfare who took a genuine interest in the developments of the project and the activities of his Residential Social Work colleagues.

Special thanks to Hilary Arlott for conscientious design of the graphic data and to Jennie Chapman for the very tolerant production of the typed manuscript.

To Dr. Alan Winton who throughout all the time and frustrations in producing this thesis supervised and continually encouraged a high standard of research practice at all times. It has been a disciplined learning experience for which I am grateful.

Finally, to Campbell, Joanna, and Helen, thanks for being so understanding of a student dad, and to Glennis for always being there.

CONTENTS

	PAGE
ABSTRACT	ii
ACKNOWLEDGEMENTS	iii
CONTENTS	iv
LIST OF FIGURES	v
LIST OF TABLES	vi
CHAPTER	
1 INTRODUCTION	1
2 METHOD	18
3 RESULTS	28
4 DISCUSSION	50
APPENDICES	
A Room Cleaning Criteria Mk I	64
Room Cleaning Criteria Mk II	65
B Room Cleaning Behaviour Recording Sheet	66
C Interobserver Reliability Checks	67
D Accuracy of Self Evaluations	
Sequence A Girls	68
Sequence B Girls	69
REFERENCES	70

LIST OF FIGURES

FIGURE		PAGE
I	Subject 1, Sequence A	29
II	Subject 2, Sequence A	30
III	Subject 3, Sequence A	31
IV	Subject 4, Sequence A	32
V	Subject 5, Sequence A	33
VI	Subject 6, Sequence B	34
VII	Subject 7, Sequence B	35
VIII	Subject 8, Sequence B	36
IX	Subject 9, Sequence B	37
X	Subject 10, Sequence B	38
XI	Mean daily scores of Sequence A combined	40
XII	Mean daily scores of Sequence B combined	41
XIII	Mean daily scores of Sequence A combined and Sequence B combined	42

LIST OF TABLES

TABLE		PAGE
1	Room Cleaning Criteria Mk III	20
2	Results Summary	39
3	Accuracy of Self Evaluations Summary	48

CHAPTER 1

INTRODUCTION

1.1 APPLIED BEHAVIOUR ANALYSIS

Applied behaviour analysis as defined by Baer, Wolf, and Risley (1968), has a number of dimensions which represent a set of principles, procedures, and a logic, for examining and intervening into the interactions between people and the environments in which they live. The principles of application are those of behaviour analysis in general, sometimes referred to as learning theory (Millenson and Leslie, 1979). The methods of application are typically those of behaviour modification and behaviour therapy (Martin and Pear, 1978).

1.2 APPLICATION TO DELINQUENT BEHAVIOUR

The use of behaviour therapy with juveniles who exhibit antisocial behaviours can be said to stem from the acceptance of an applied behaviour analysis approach to delinquency. Parsonson (1972) has given a rationale for such an approach, and has suggested methods by which treatment of delinquent behaviours can be effected. Parsonson (1972, p.50),

Delinquents behave as they do because they lack certain pro-social responses - either because these responses have not been acquired or because of lack of reinforcement, they have either been extinguished or have a low probability of emission. They generally conform to a different set of norms, because behaviours appropriate to those norms have positive reinforcing consequences. Attempts to modify delinquent or anti-social behaviours

must, therefore, be based upon some programme which reduces the reinforcing properties of socially deviant responses and increases the probability of emission of socially accepted behaviours by reinforcing them. The application of the laws of extinction and reinforcement, derived from learning theory, seems to match the above requirement most closely, thus behaviour modification appears as a treatment of choice.

It was an acceptance of the principles of applied behaviour analysis, and the application of them to a treatment programme for delinquent girls in residential care that has resulted in the reporting of the development of self monitoring procedures contained in this thesis.

Applied behaviour analysis programmes have been concerned with delinquent behaviour in many and varied settings, for example: correctional institutions (Ayllon and Milan, 1979); community based group homes (Fixsen, Phillips and Wolf, 1978); juvenile halls (Allison, Kendall and Sloane, 1979); probation services (Tharp and Wetzel, 1969); law enforcement agencies (Schnelle, Kirchner, McNeese, and Lawler, 1975); youth centers (Pierce and Risley, 1974); social work settings (Schwartz and Goldiamond, 1975); and situations calling for crime prevention (McNeese, Egli, Marshall, Schnelle and Risley, 1976). Reviews of this literature have been done by Braukmann and Fixsen (1975); Braukmann, Fixsen, Phillips and Wolf (1975); Burchard and Harig (1976); Davidson and Seidman (1974); Johnson (1977); Milan and McKee (1974); Morris (1978); and Stumphauzer (1973, 1979).

1.3 CURRENT ISSUES IN APPLYING BEHAVIOUR ANALYSIS TO DELINQUENCY

Accompanying the wider use of behaviour analysis techniques and procedures for treating delinquent behaviours, has been an increasing focus on the utility of the approach. A brief consideration of these issues follows.

1.3.1 Social Validity

Wolf's (1978) article on social validity implied that behaviour analysts must be sensitive to how the goals they select fit with those of the social systems in which they work. This is a particularly important issue in the treatment of delinquency as the social significance of behavioural goals needs to be evaluated from the perspective of different groups of relevant judges from the community which would include, general community members, school officials, social workers, law enforcement officers, magistrates, correctional personnel, or delinquents and criminals. Social validity measures can be misleading, manipulated and abused thus care is needed in sampling for relevant judges. Social validation has so far been applied in treatment programmes for delinquent youth by Braukmann et al., (1975), Minkin et al., (1976), Werner et al., (1975) and Willner et al., (1977), with some encouraging results.

1.3.2 Methodology

Reliable measurement and evaluation of delinquent behaviours is a major problem. They occur infrequently in the lives of most people, even those labelled delinquent or criminal, and are not always easily detected at the time

they occur. Because of this, programmes of direct analysis and modification can rarely be applied, with the result that an indirect approach - one that attacks behaviours related to the presenting problem must be used (Baer, 1975). Emery and Marholin (1977) suggested that delinquent behaviours per se might be directly suppressed, but concluded that trying to punish directly would be extremely difficult because delinquent behaviours often occur under covert stimulus conditions over which little control can be achieved.

The alternative is to target behaviours believed to be incompatible with delinquency (Bassett, Blanchard and Koshland, 1975). These include social/interpersonal, vocational, programme, compliance and school behaviours. Hence, if desirable skills of gaining relevant reinforcers by other than illegal means are taught well enough, then no direct suppression of delinquent behaviours is necessary. Baer (1975) has suggested that the undesirable responses may simply fall into disuse, displaced by the desirable (and more profitable) skills. Baer (1975) has proposed a model that could demonstrate this process empirically and the Achievement Place Research Group analysed delinquency measures subsequent to the modification of specific target behaviours (Braukmann, Kirigin and Wolf, 1976). Although not complete, data suggest that a concurrent decrease in measures of delinquency does occur (Willner et al., 1978).

1.3.3. Generalization and Maintenance

Generalization involves the transfer of treatment from one setting to another and maintenance implies durability

over time. These are two crucial issues for applied behaviour analysis programmes involved with treating delinquents. An expectation of behaviour change not only occurring during treatment but maintaining after the cessation of treatment is commonly associated with these programmes. Some of the strongest criticisms of behaviour analysis work, particularly in the area of delinquency, have been on lack of generalization and maintenance (Johnson, 1977; McCombs, Filipczak, Friedman and Wodarski, 1978; Beck and DeVoge, 1978; Davidson and Wolfred, 1977; Emery and Marholin, 1977). From the literature in the area, it seems essential that generalization must be planned for and a number of specific techniques have been suggested for incorporation into treatment programmes (Kazdin, 1972; Marholin, Siegel and Phillips, 1976; Stokes and Baer, 1977), in order to enhance generalisation outside of the treatment settings. This thesis is particularly concerned with the development of a self evaluation procedure to assist maintenance of treatment gains outside of an institution setting.

1.4 THE TOKEN ECONOMY

The token reinforcement system introduced by Ayllon and Azrin (1968) has been one of the most widely used, and most adaptable procedures for the application of applied behaviour analysis to treatment programmes. This system has been demonstrated to be effective and useful for modifying and maintaining various behaviours in a wide variety of settings by employing the use of conditioned generalized reinforcers. Appropriate behaviour is rewarded with tokens

which are later exchanged for specified privileges. Simultaneously, inappropriate behaviour is either punished by the fining of tokens, the reinforcement of an incompatible desirable behaviour, and/or by extinction of the inappropriate behaviour. Johnson (1977) summarised 29 studies from 1965 to 1976 using token economy procedures in what he classified as correctional institutions. Of these, 18 were with delinquent and predelinquent youth, three with delinquent soldiers, two with antisocial retardates, four with adult offenders, one with psychopathic youth and one in a maximum security hospital. The general preference for use with youthful offenders is apparent. The majority of behaviours shaped were classified as - institutional convenient, educational, or appropriate social, - in other words, clear examples of an indirect approach that attacks behaviours related to the presenting problem, described by Baer (1975), and which are considered incompatible with delinquency. None of the programmes referred to that employed a token economy for positive reinforcement reported a failure to obtain an increase in the target behaviour.

1.5 THE ADDITION OF SELF CONTROL PROCEDURES TO TOKEN PROGRAMMES

It is now accepted that there exist numerous instances of token programmes producing behaviour changes, only while contingent reinforcement is being delivered. Kazdin (1972) and Stokes and Baer (1977) have gone further, to describe a number of methods that could be used to enhance generalization outside of the treatment setting. In view of the fact that maintenance of behaviour change

typically deteriorates when token programmes are withdrawn, some reviewers have suggested that self regulation and self evaluation procedures be incorporated into token programmes (Bandura and Perloff, 1967; Kazdin and Bootzin, 1972; O'Leary and Drabman, 1971).

1.5.1 It appears that used as an isolated procedure, self assessment has not been particularly effective. Both Santogrossi, O'Leary, Romanczyk and Kaufman (1973, and Turkewitz, O'Leary and Ironsmith (1975) instructed disruptive children to make global ratings of their social behaviour in the classroom. In neither study was there a reduction in disruptive behaviour. Similarly a study by Layne, Rickard, Jones and Lyman (1976) demonstrated the failure of self-assessment to strengthen room cleaning behaviour. On the other hand, Sagotsky et al., (1978) found a self-assessment procedure to be highly effective in reducing off-task behaviour and increasing academic rate and accuracy, and Nelson, Lipinski, and Boykin (1978) successfully taught retarded adolescents to increase their rates of appropriate verbalizations by using a self recording procedure.

1.5.2 Three factors appear to influence the effectiveness of self assessment procedures: i) accuracy of assessments; ii) difficulty of the task; iii) the type of person.

Accuracy of Assessments: While accurate assessment is possible (Santagrossi et al., 1973), reliability with independent observers needs additional training. Hundert and Bucher (1978) and Nelson et al., (1978) successfully taught increased accuracy of self recording. The training

procedures involved contingently rewarding recordings which matched those of independent observers. Matching contingencies can be faded to a minimal checking routine without loss of accuracy (Wood and Flynn, 1978). Despite this, in neither case did increased accuracy lead to better performance.

Difficulty of the Task: Peacock, Lyman and Rickard (1978) hypothesised that the effect of accuracy on performance might depend on task difficulty, and they examined the reactivity of self monitoring room cleaning tasks. When adolescent boys were rewarded for accurately monitoring easy tasks, both accuracy and task performance increased. In contrast, increased reliability of monitoring hard tasks was not accompanied by parallel task improvements. Thus accuracy may be a factor in determining the reactivity of self-assessment when the task is easy.

The Type of Person: It is interesting to note that the studies by Santagrossi et al., (1973), Turkewitz et al., (1975) and Layne et al., (1976) involved children with a history of behavioural difficulties who possibly did not generate negative self evaluations when rating their behaviour as clearly inappropriate. A study by Sagotsky et al., (1978) used unselected subjects and was suggestive that self assessment might be a useful procedure for persons who demonstrate a clear motivation to improve their behaviour.

Other support for a motivational factor comes from evidence that self assessment may not be effective initially, but may be successful in maintaining behaviour change achieved via an externally imposed token programme. Bolstad and Johnson (1972), Turkewitz et al., (1975) and Seymour and

Stokes (1976) implemented a system of self assessments without rewards, following successful token programmes. In all cases the effect of the token programmes were maintained with self assessment alone, possibly because the token programmes increased the motivation to exhibit appropriate behaviour, or increased the likelihood that self evaluations prompted covert self reinforcing statements.

1.5.3 In a number of other studies, self assessment has been combined with reinforcement procedures, and its additive effect has been evaluated. Salzberg (1972) found no incremental effect of self assessment over a strong contingency. Knapczyk and Livingston (1973) found that a system of tokens plus self assessment was no different from tokens alone in terms of effects on accuracy. Only Seymour and Stokes (1976) reported increments in the amount of work behaviour, when self assessment was added to a relatively ineffective token programme. It may be that additive effects of self assessment are obscured by powerful reward programmes.

1.5.4 Comparative research indicates that self assessment is just as effective as external assessment when both are followed by rewards. Bolstad and Johnson (1972) compared self and external procedures and found no differences between self and external assessments. Frederiksen and Frederiksen (1975) demonstrated a similar result, as did Wood and Flynn (1978).

1.5.5 The usefulness of self assessment as a maintenance strategy has been demonstrated. However, the question of

whether experience using self evaluation facilitates maintenance when both assessment contingencies are withdrawn has been addressed by Wood and Flynn (1968) and results were encouraging. The self evaluation group maintained high levels of task performance throughout the study. The external evaluation group showed a marked decrease, at a subsequent baseline. The hypothesis that self evaluation would facilitate maintenance following programme termination was confirmed.

1.5.6 Self assessment used alone does not appear to effect significant changes in behaviour where clinical interventions are required. It may be more effective with persons already motivated to perform appropriately. Accuracy of the assessments and task difficulty are also apparently important factors to consider. Self assessment does not add significantly to the effects of an already effective reward system. When used in conjunction with rewards self assessment is as effective as external assessment. Self assessment may be useful for maintaining effects when other interventions are withdrawn. Experience using self assessment facilitates maintenance of treatment gains even when all programatic treatment is terminated.

1.6 BEHAVIOUR PROGRAMME IN OPERATION

1.6.1 Prior to this study a behaviour programme had been developed and was in operation at the Home. This consisted of a behaviour report, covering nine specific areas of behaviour: Dormitory, Dining Room, Work, Hygiene, Boundary Keeping, Verbal, Non-Verbal, Compliance and

Temperament. Girls could earn or lose points for appropriate or inappropriate behaviour in these areas. Criteria were established for each of these areas, but they were quite broadly interpreted by the staff. A high degree of tolerance to deviations from the criteria occurred. The reports were marked three times daily, after lunch for the morning, after tea for the afternoon, and at bedtime for the evening. A girl could earn up to three credits per behaviour area each day. These credits translated to pocket money earnings, which were provided weekly.

1.6.2 The entire programme was based on external monitoring of a girl's behaviour by staff. When inappropriate behaviour occurred, the staff member would intervene and inform the girl that the behaviour was unacceptable and remind her of the appropriate behaviour required. In effect the staff acted as teachers of appropriate behaviour, similar to the role of Teaching Parents at Achievement Place (Fixsen, Phillips and Wolf, 1978). Despite the lack of technical precision, the reports provided the girls with a measure of their behaviour in each category.

Although the programme in operation has its shortcomings, girls did learn more appropriate methods of behaving, and generally improved the overall standard of their personal and interpersonal behaviour, and social skills whilst in residence.

1.6.3 During 1979 a number of doubts began to be raised about the efficacy of a totally externally-monitored behaviour programme. The literature on the topic of

institutional behaviour modification programmes has illustrated quite clearly that changes in behaviour within a residential setting are relatively easy to effect by the use of token economy procedures (Kazdin, 1972). The major problem however, is how to have these changes generalize outside of the institution setting (Stokes and Baer, 1977).

1.6.4 The incorporation of self-monitoring procedures into token programmes has produced promising results, for assisting generalization and maintenance of appropriate levels of behaviour beyond residential token economy settings (Wood and Flynn, 1978).

1.6.5 The presenting problems seen in the Girls' Home were twofold. First, there was a general dependence by the girls on the staff's external monitoring of the behaviour reports for the earning of pocket money and privileges. Secondly, there was a general reluctance by girls to take full responsibility for their own behaviour. Too often staff members took on the responsibility for monitoring the girls' behaviour while they were in residential care. A major problem arose at discharge when a girl was rather abruptly weaned off the behaviour programme and expected to maintain a high standard of acceptable behaviour after returning to the community.

1.6.6. A major goal for the home was to try and establish greater personal responsibility by the girls for behaviour whilst in residential care in order to produce more self control over their own behaviour, and the capability to maintain this self control at an acceptable level when

they returned to the community.

The present study was instigated as an initial step.

1.6.7 The study undertaken was essentially a programme within a programme as the normal residential token economy and behaviour reports of the home operated as usual.

Dormitory was one of the nine specific areas of behaviour included in the residential token economy, for which up to 3 points per day could be earned. Points were awarded on the basis of one per time period - morning, afternoon and evening - providing the criteria were met.

The criteria for Dormitory Behaviour was:

Morning:

- 1) Rise when called.
- 2) Beds stripped and made up neatly with bedclothes tucked in.
- 3) All spare clothes folded and placed in wardrobes or drawers.
- 4) Rooms to be left neat and tidy prior to leaving.

Afternoon:

- 1) On return from school or work, when clothes are changed, uniforms etc., to be hung in wardrobes.
- 2) Casual clothes changed into.
- 3) Room to be left neat and tidy prior to leaving.

Evening:

- 1) All clothes to be folded and put away.
- 2) Clothes for next day to be ready for use.
- 3) Room to be neat and tidy prior to getting into bed.

During Baselines I and II and Instructions phases girls could still receive up to three points per day for

Dormitory Behaviour from the residential token economy.

During Baseline III, External Tokens, and the three Self Evaluation phases, the three points per day for Dormitory Behaviour from the residential token economy were withdrawn.

1.7 ROOM CLEANING SELF EVALUATION EXPERIMENTAL STUDY

1.7.1 The goal of this study was to develop an accurate behavioural self evaluation and reporting procedure which established and maintained appropriate levels of behaviour, with institutionalised delinquent girls. The basic objectives were to establish for each girl accurate self evaluation, establishment and maintenance of appropriate room cleaning behaviour.

1.7.2 Essentially the study was designed as a systematic replication of Wood and Flynn (1978) whose goals were to determine the effects of self evaluation, as compared to an external token system, and to demonstrate a method of establishing accurate self evaluation and management of appropriate behaviour in pre-delinquent youths. Their objectives were to establish for each youth accurate self evaluation of room cleaning behaviour and to establish and maintain appropriate room cleaning behaviour.

1.7.3 The objective of determining the effects of self evaluation as compared with an external token system of Wood and Flynn, although not a specific goal of this study was still incorporated into the replication design.

1.8 FACTORS AFFECTING THE SYSTEMATIC REPLICATION

1.8.1 Setting

The Living and Learning Centre in Fort Lauderdale in Florida was a family style residential rehabilitation programme caring for youths who had been placed by Florida State Agencies or committed by the Juvenile Courts, basically the size and function of a Department of Social Welfare Family Home in this country. This systematic replication was done within a Department of Social Welfare Short Term training institution, caring for girls who had been admitted to the Home, either by Social Workers, members of the Police, or as a result of Children and Young Persons Court proceedings.

1.8.2 Staff

The Living and Learning centre was staffed by professional teaching parents who had received training and certification from the Achievement Place programme at the University of Kansas. The staff of the Girls Home were Residential Social Workers, employed by the Department of Social Welfare, who had a basic working knowledge of behavioural techniques used in residential care, but without any specific formal training. The ratio of staff to residents at the Living and Learning Centre was two to six, and at the Girls Home was usually two to fifteen.

1.8.3 Subjects

The average age of youths at the Living and Learning Centre was thirteen years, and the length of stay averaged 13 months at the time of the study. The average age of girls at the Girls Home was fourteen years, and the length

of stay averaged one week at the commencement of the study. Affecting stability was a continual turnover of relatively short stay residents at the Girls Home, whilst the Living and Learning centre's six residents remained stable throughout that study. The youths at the Living and Learning centre had separate rooms, whereas the girls at the Girls Home all shared rooms, with a minimum of two and up to four in each.

1.8.4 Programme

Both Institutions had programmes based on a Token Economy. There was similarity in the selection of behaviours incorporated. The points in the Token Economy of the Living and Learning centre were used internally for the purchase of privileges, with limited use for purchasing activities outside of the residential setting. The points earned at the Girls Home converted to pocket money at a cent a point, and this money was used exclusively outside of the institution by the girls themselves. All privileges were easily accessible within the home and were available on a non-contingent basis.

1.8.5 Summary

In a number of respects this systematic replication was conducted under quite different circumstances to those of the Wood and Flynn study.

Perhaps the major constraint experienced at the Girls' Home was the time available to conduct the study. In contrast Wood and Flynn had a stable population for a long period of time (up to 19 months). This study had a

continually changing transient population whose maximum period of residence was up to three months. It was considered necessary though, to attempt the study within the shorter time scale as most residential programmes in this country are relatively short term, and if the development of self monitoring is to be viable it would need to be demonstrated within the usual length of time children and young persons are in residential care.

Despite all the factors affecting the replication every effort was made to maintain the technical precision required for the definition of the target behaviour, the reliability of the observations and the treatment procedures.

CHAPTER 2

METHOD

2.1 SUBJECTS AND SETTING

The subjects in this study were 28 girls admitted to the Department of Social Welfare Girls' Home in Palmerston North, between August and November 1980. The ages ranged from 12 to 16 years, with a median age of 14 years. Approximately one third (9) of the girls were State Wards and two thirds (19) were on remand or adjournment from the Children and Young Persons Court.

The Girls' Home was established primarily to provide a short-term training programme for girls aged between 10 and 17 years, with the usual period of residence approximately one school term (i.e. up to three months). The Home also produced assessments on all admissions, regular monthly progress reports and attendant recommendations for placement on discharge.

The length of stay for the girls involved in the study varied between one and four months. Thus very few girls were resident in the home throughout all phases of the study.

The Pilot section was instituted during the final two weeks of the second term when the criteria were finalised, and the study proper commenced with new girls admitted during the early stages of term three. Some girls admitted on very

short adjournments (less than one month) had data recorded to provide uniformity of treatment for all girls in residence at the time, but these data have not been included. The relatively high turnover of girls in the home made it difficult to obtain sufficient data from individual girls throughout all phases of the study.

Despite these difficulties, 10 subjects provided sufficient data for analysis. Eighteen subjects were excluded, 9 due to completing only one phase before discharge, 6 due to completing only one baseline and one treatment phase (with no reversal, in effect a quasi-experimental design), before discharge, and 3 were excluded due to not completing either the external tokens or Self Evaluation I phase.

2.2. RECORDING AND RELIABILITY

2.2.1 Target Behaviours

Room-cleaning behaviour was the dependent variable during this study. Ten behavioural definitions were modified from Fixsen et al., (1972), Weir and Ford (1977), and Wood and Flynn (1978) specifying the degree of cleanliness or the placement of objects in the girls' bedrooms (see Table 1 for definitions). Each bedroom was shared by a minimum of two and up to a maximum of four girls.

2.2.2 Observations

The Residential Social Work staff recorded whether or not each area of a girl's room met each of the 10 room cleaning criteria. Room cleaning was assessed and recorded

TABLE I

ROOM CLEANING CRITERIA

MK III

1. BED - Pillow in centre of bed touching headboard, open end of pillow away from door. Bedspread tucked in at foot of bed, sides hanging down evenly, folded completely over pillow, with fold line visible under pillow. Quilt placed on top of bed not overhanging the sides.
2. BED - Made up with 1 under blanket, 2 sheets, top blanket(s), 1 bedspread, 1 quilt. Sheets and blankets tucked in all the way around bed, top sheet folded over 1 pillow width from bedhead.
3. WARDROBE CLOTHING - All non-folded clothes to be hung on hangers, all hangers to be on rail, hooks facing rear wall of wardrobe.
4. WARDROBE FOOTWEAR - All footwear to be placed together in pairs, with toes touching rear wall of wardrobe.
5. BEDSIDE CABINETS, SETS OF DRAWERS INTERIOR - All clean clothes folded, other personal belongings sorted in an orderly fashion.
6. BEDSIDE CABINETS, SETS OF DRAWERS EXTERIOR - Tops to be free of dust, no upright objects to overhang any edge. Cabinets and drawers no further than skirting board width from any wall.
7. POSTERS, PICTURES - Can only be pinned onto either the wall hangings or the display boards. No drawing pins to be used.
8. CURTAINS - To be hanging from tracks attached to runners, drawn back during day to their fullest extent. Interior windows, sills and frames to be free of dust and dirt.
9. FLOORS - To be free of any dust and dirt, mirrors to be free of dust and dirt.
10. LAUNDRY BAGS - Laundry bags containing dirty clothing to be emptied into basket each morning, then hung from hook inside wardrobe. No clothing or personal belongings on beds, floors, or furniture, apart from personal objects on bedside cabinets and sets of drawers.

each day by the morning shift staff, usually after most girls had gone to school. During Baselines, the girls were never told that their rooms were being checked, although some of the girls remaining in the residence did observe increased activity by staff in the dormitory areas.

At least once during each experimental phase, the afternoon shift staff recorded room cleanliness, independent of the morning shift staff. The two sets of recordings were compared and scored, item by item, for the number of agreements for each girl's room.

As a reliability check the percentage of inter-observer agreement was obtained by dividing the number of agreements by the total number of agreements plus disagreements and multiplying by 100.

Six reliability checks were made during Baseline I, two during Instructions, two during External Tokens and Self Evaluation I, one during Baseline II, one during Self Evaluation II, one during Baseline III, and one during Self Evaluation III. For the fourteen reliability checks made, inter-observer agreement ranged from 78.8 percent to 93.6 percent with a mean of 87.4 percent.

2.3 PROCEDURES

2.3.1 Design Elements and Experimental Conditions

The study consisted of several experimental conditions during which room cleaning behaviour was recorded each morning of the week.

2.3.2 Baselines I and II

During these phases there were no scheduled consequences for room cleaning. The observers merely recorded the cleanliness of each girl's room according to the ten criteria. The normal residential token economy for dormitory behaviour kept functioning with three points per day being available (criteria outlined in Introduction).

2.3.3 Baseline III

This was identical to Baselines I and II except that the residential token economy with three points per day available was withdrawn.

2.3.4 Instructions

This phase was also identical to Baselines I and II except that, in addition, instructions as to the behaviours comprising room cleanliness were given. These instructions were verbally explained, demonstrated and practiced by each girl. A written copy of the criteria was posted behind each wardrobe door.

2.3.5 External Token System

Under this condition, a girl's room cleaning behaviour received externally determined consequences. The Residential Social Worker evaluated the girl's room cleaning behaviour scoring from 0 - 10 points per day, one point per correct item. The evaluation was made prior to the girl going to breakfast, when room cleaning was expected to be completed. Immediately following the assessment, feedback as to why some areas received points and others did not was given to the girl. During this

phase the usual residential token economy for room cleaning of three points was withdrawn.

2.3.6 Self Evaluation Token System

2.3.7 Self Evaluation I Phase

Under this condition a girl evaluated the cleanliness of her own room using the same list of ten definitions as the staff observers (Table 1). She received two types of points:

- a) Points related to the degree of room cleanliness;
- b) Points related to accuracy of evaluation.

To ensure that the maximum daily point earnings did not exceed the Sequence A girls who were concurrently being externally evaluated, the Sequence B girls received a $\frac{1}{2}$ point for each item they scored as completed according to the criteria, (a maximum of 5 points) and a $\frac{1}{2}$ point for the evaluation of each item that matched the Residential Social Worker's evaluation of that item (a maximum of 5 points). This meant a maximum of ten points per day was available. The residential token economy of three points per day was withdrawn during this phase.

Accuracy points were determined by the number of agreements, item by item, between the girls own evaluation and that of the Residential Social Worker. This system was identical to that used for the inter-observer reliability checks, only instead of comparing reliability between staff and staff, it compared reliability between girl and staff.

During this phase, daily matching for accuracy was

done for the first seven days to ensure that a minimum of 80 percent accuracy was achieved. After accuracy had been maintained, a girl was told that her evaluations would be spot checked on a random basis, and if her evaluation failed to match the Residential Social Worker's evaluation at the 80 percent level, she would lose accuracy points, and would return to daily matching. If accuracy was maintained through the initial stages of self evaluation and later during the spot checks, then the matching procedure was eliminated completely. Following the 7 days of matching one point per item scored correct by the girl was instituted. The return to matching was not required as the 80 percent level of accuracy was maintained.

2.3.8 Self Evaluation II Phase

This was identical to the Self Evaluation I phase except that there was no daily matching for accuracy unless an evaluation fell below 80 percent accuracy during a spot check.

2.3.9 Self Evaluation III Phase

There was no scheduled reinforcement during this time but girls continued daily self evaluations of their rooms.

In all three self evaluation phases each girl determined the eventual point consequence for her performance, apart from the matching phase, when the allocation of points was moderated by the Residential Social Worker. In contrast, during the External Token System phase, the evaluation and allocation of consequences were fully controlled by the Residential Social Worker.

2.3.10 Treatment Group and Sequence

The girls were randomly assigned to the two treatment groups, Sequence A and Sequence B. Of the ten girls whose data were analysed, five received Sequence A and five received Sequence B. Some gaps in continuity were inevitable due to scheduled and unscheduled absences from the home. Any girls admitted to the home after the start of the programme participated in the particular phase running at the time of their admission.

1. Sequence A group received: Baseline I (15 days), Instructions (15 days), External Tokens (13 days), Baseline II (13 days), Self Evaluation II (12 days), Baseline III (7 days), and Self Evaluation III (7 days).
2. Sequence B group received: Baseline I (15 days), Instructions (15 days), Self Evaluation I (13 days), Baseline II (13 days), Self Evaluation II (12 days), Baseline III (7 days), Self Evaluation III (7 days).

Equivalent phases in Sequence A and B ran concurrently.

The five girls in Sequence A received both the External tokens and Self Evaluation II and III (with random checks) to allow comparison among themselves of their performance under both treatments.

The five girls in Sequence B received the Self Evaluation condition three times; once with accuracy matching, and twice with random checks to allow comparison amongst themselves. Sequence B also allowed comparison with Sequence A during the External Token condition.

2.3.11 Reinforcement Considerations

The back-up reinforcement in this experiment was money. Each point was worth one cent and was given in addition to the pocket money earnings provided by the normal token economy system operating within the home. This money could be used either for outings to town, or weekends at home. The minimum criterion for unsupervised town outings was earning \$2 per week. Generally under the normal token economy system alone, less than one quarter of girls in residence earned sufficient to obtain this privilege each week. The probability of reaching the criterion for an outing could be considerably boosted by the points earned for room cleaning. If a girl performed well on the room cleaning experiment, as well as keeping up a high standard in the normal token programme, she could collect a bonus on her earnings which she could take with her on weekend leave.

Both types of points proved to be strong reinforcers for the girls in residence, increasing the behaviours desired under each condition, when they were given as consequences. In effect, this study provided a bonus system in addition to the normal token economy programme.

As previously mentioned, reductions in available reinforcement were also instituted during the course of the experiment. Higher levels of reinforcement were provided initially, but were reduced towards the terminal stages.

1. During Baselines I and II, and Instructions,

the normal token programme was functioning with three points per day being available.

2. During External Tokens ten points per day were available, 1 per item correct. The normal token programme was withdrawn.
3. During Self Evaluation I and II 10 points per day were available. The normal token programme was withdrawn.
 - Contingencies during matching were a $\frac{1}{2}$ point per item completed and a $\frac{1}{2}$ point for matching the Residential Social Worker's evaluation accurately.
 - Contingencies when matching was withdrawn were 1 point per item completed according to a girl's self evaluation.
4. In Baseline III and Self Evaluation III the normal token programme was withdrawn and no points were available either from this or for room cleaning.

CHAPTER 3

RESULTS

3.1 GRAPHS

3.1.1 Individual Data

Figures 1 to 5 show the individual performances for the five girls who received Sequence A (External Tokens, and Self Evaluation phases). Figures 6 to 10 show the individual performances of the five girls who received Sequence B (Self Evaluation phases only). Missing data points on the graphs occurred due to illness, weekends at home, court appearances out of town, and absences whilst missing from the institution.

Each data point represents the individual score (out of a possible ten) for room cleaning behaviour on each recorded day. The open circles represent the data points and are:

- i) the staff member's scores during Baselines I, II and II, Instructions, and External Tokens phases;
- ii) each girl's own self-evaluated score during Self Evaluation I, II and III phases.

The crosses represent the check points, scores given by the staff member during the matching phases of Self Evaluation I, and during the random spot checks in the Self Evaluation II and III phases.

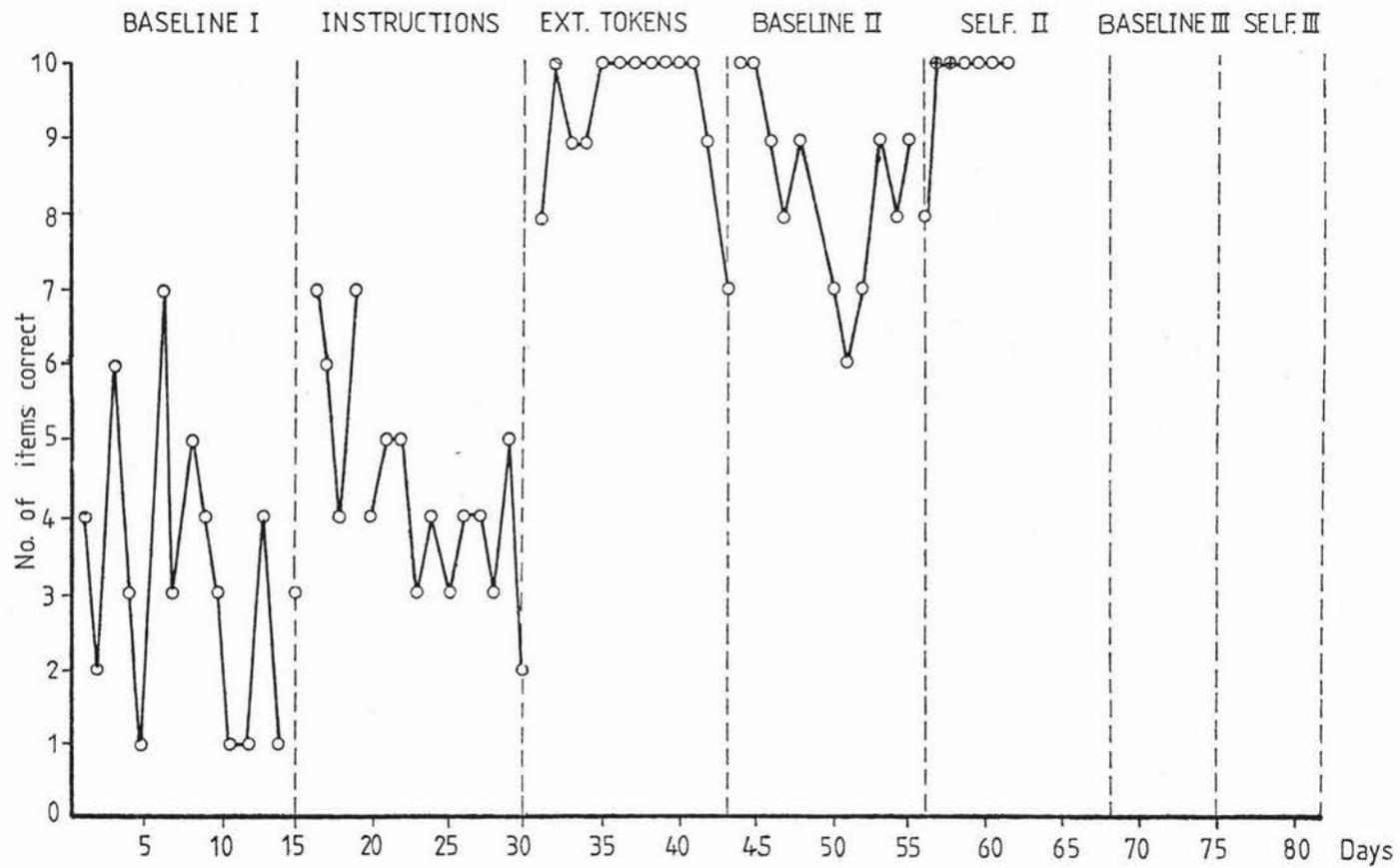


Fig.I. Subject 1, Sequence A

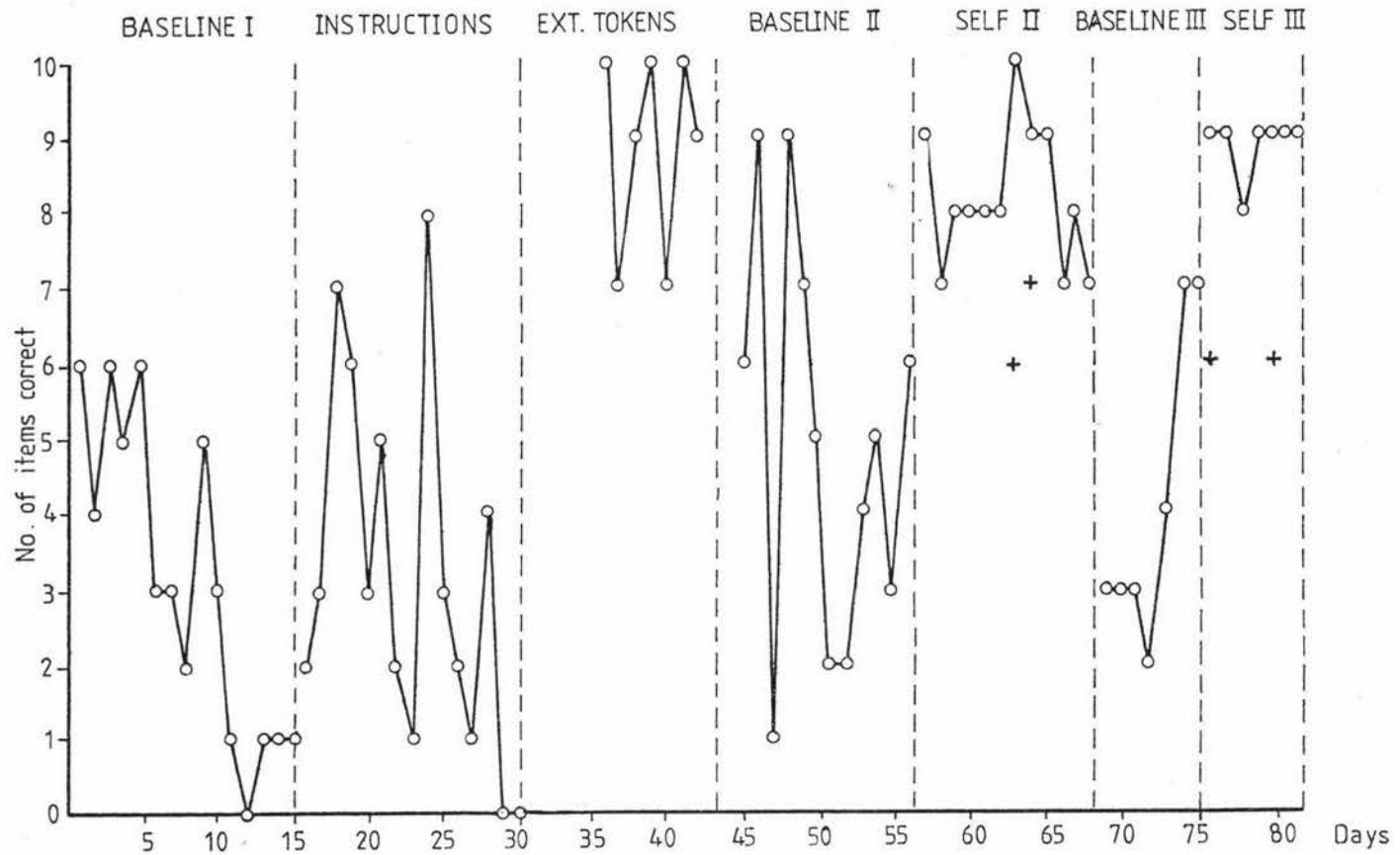


Fig. II. Subject 2, Sequence A

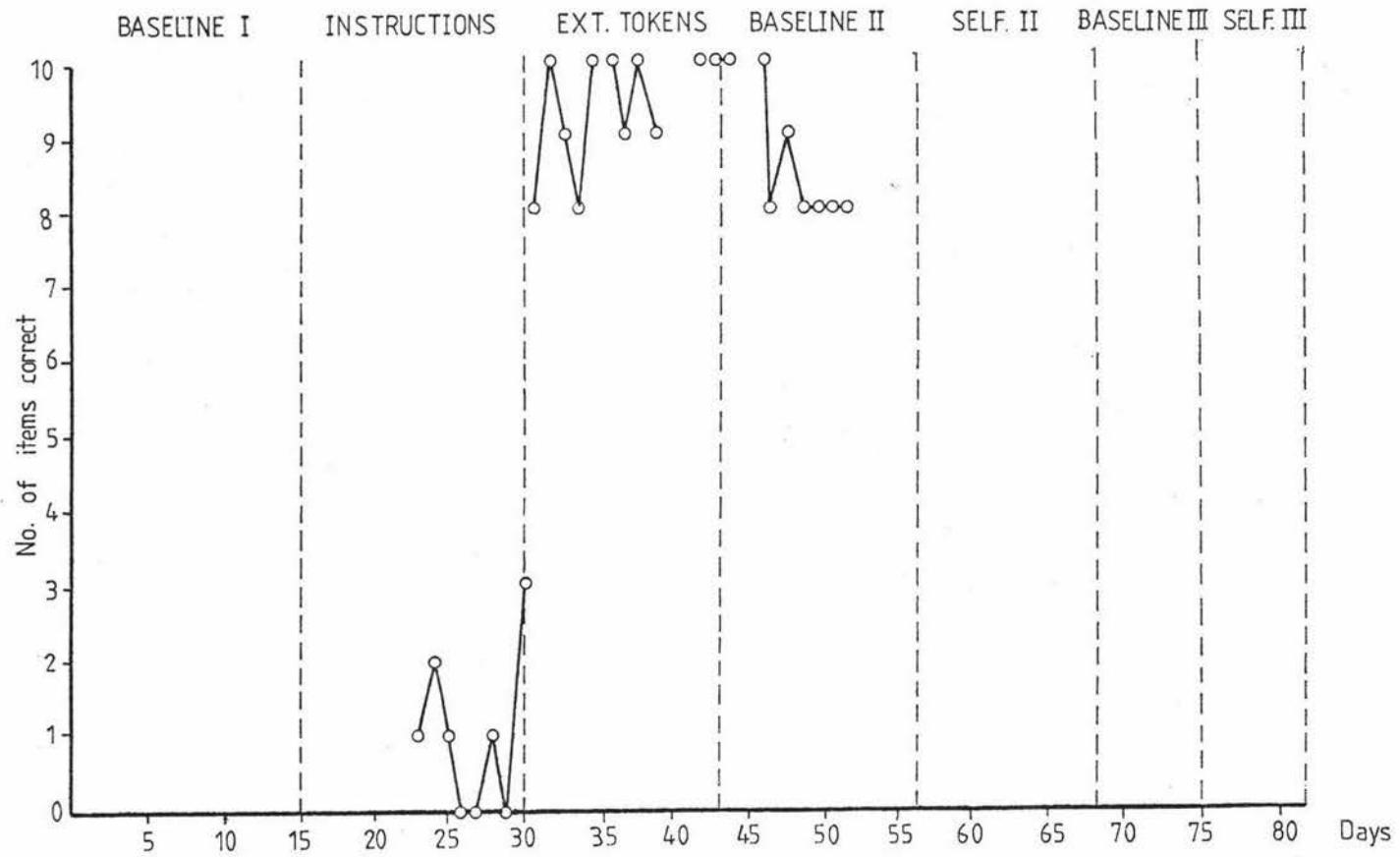


Fig. III. Subject 3, Sequence A

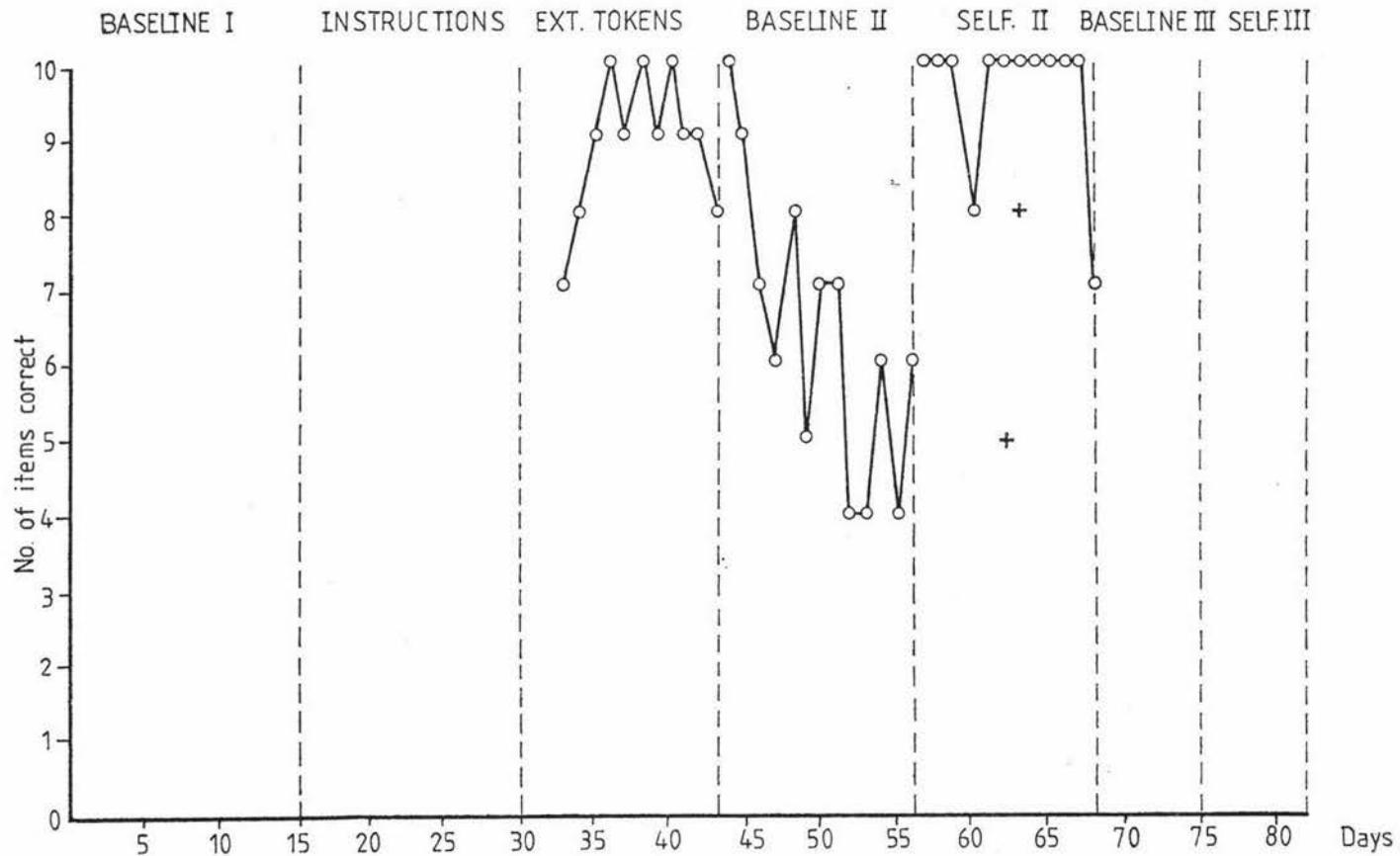


Fig. IV. Subject 4, Sequence A

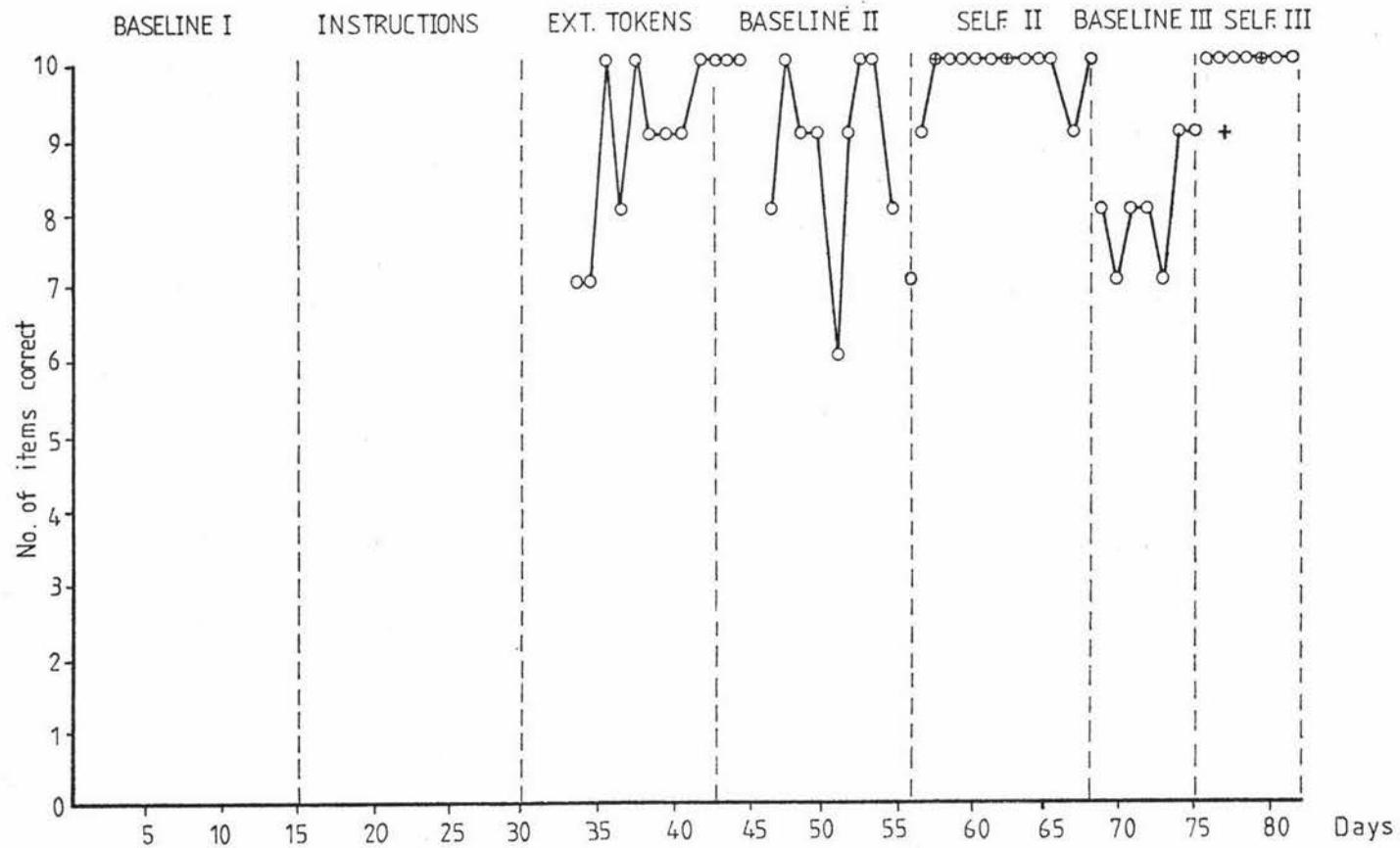


Fig.V. Subject 5, Sequence A

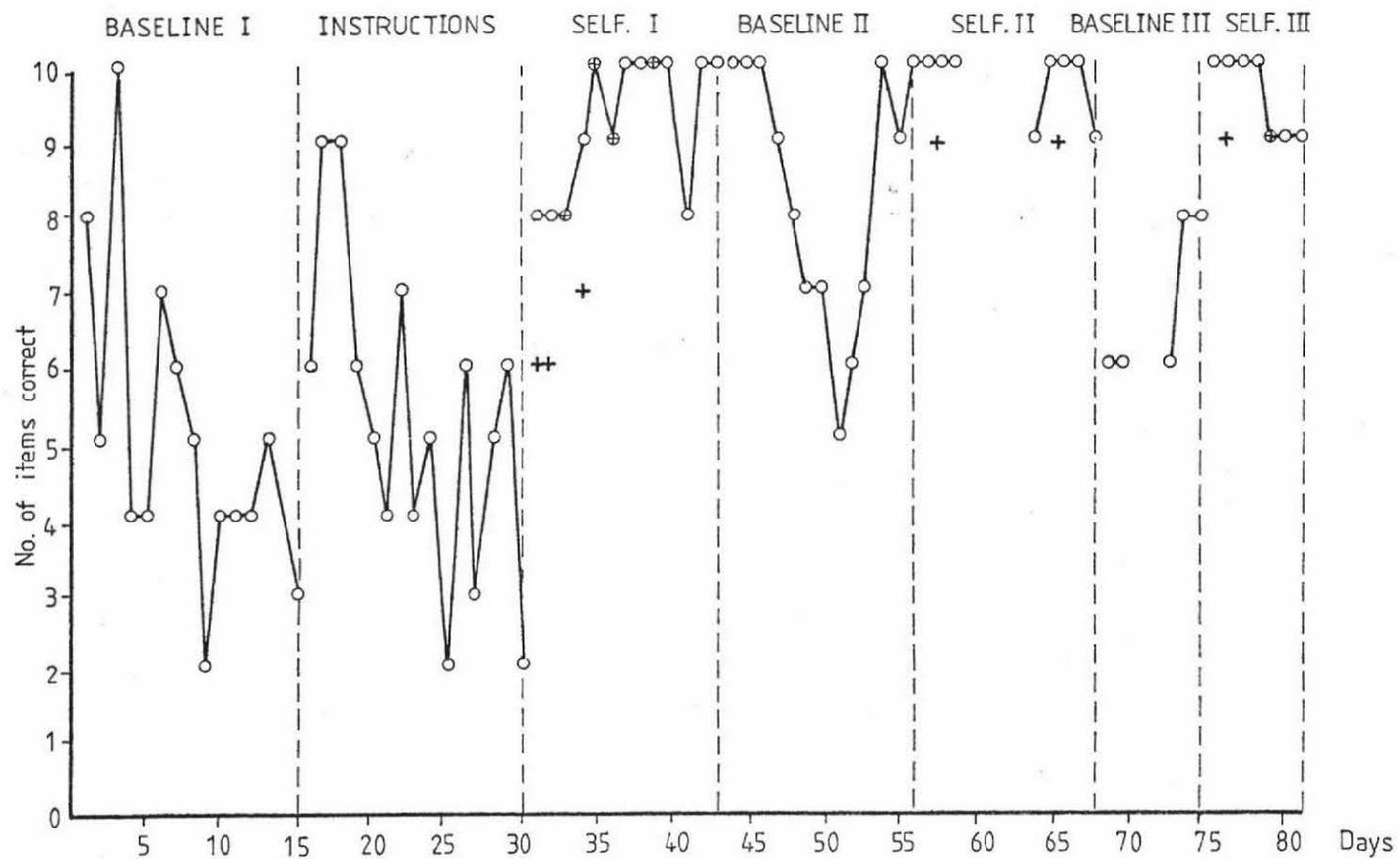


Fig. VI. Subject 6, Sequence B

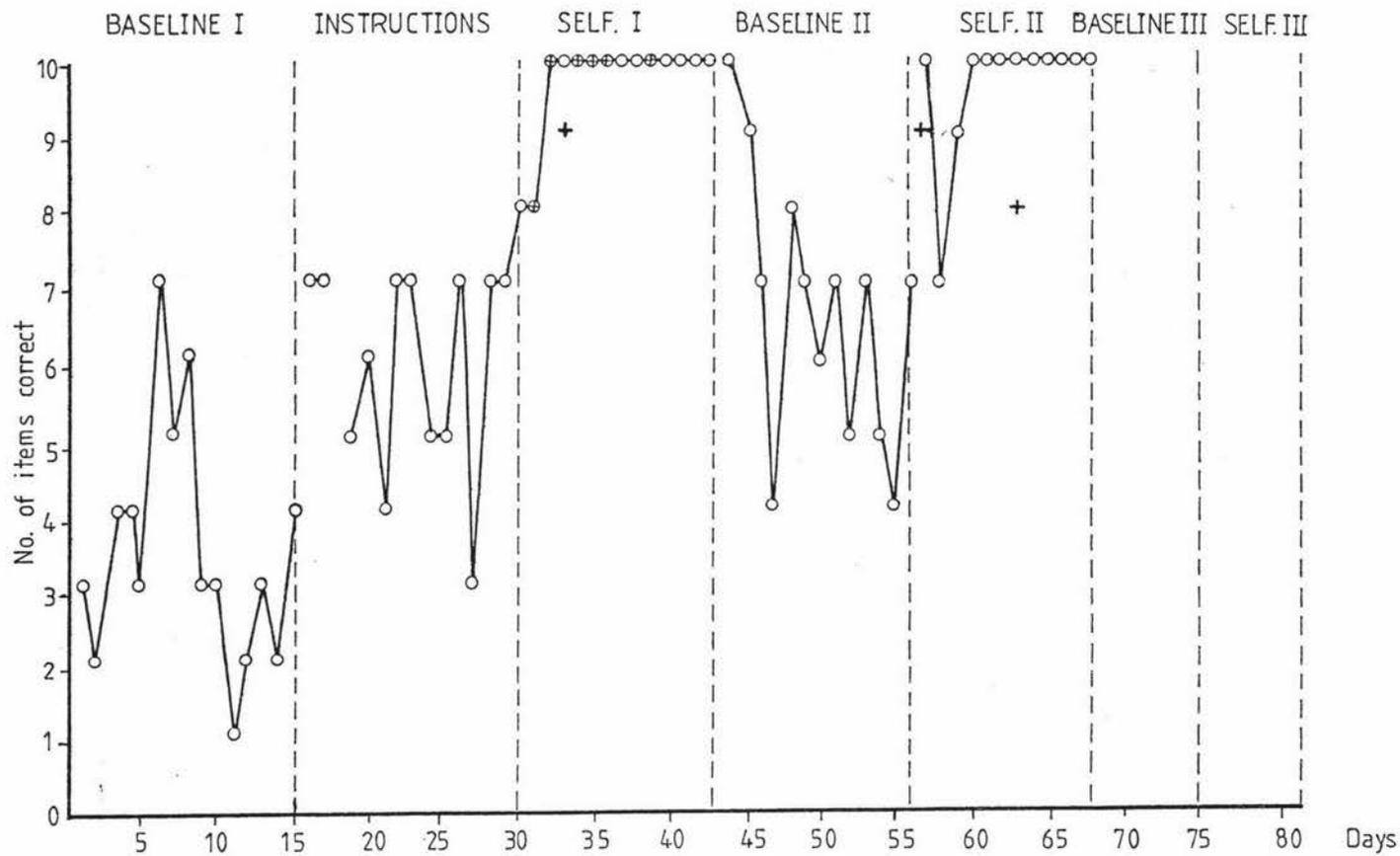


Fig.VII. Subject 7, Sequence B

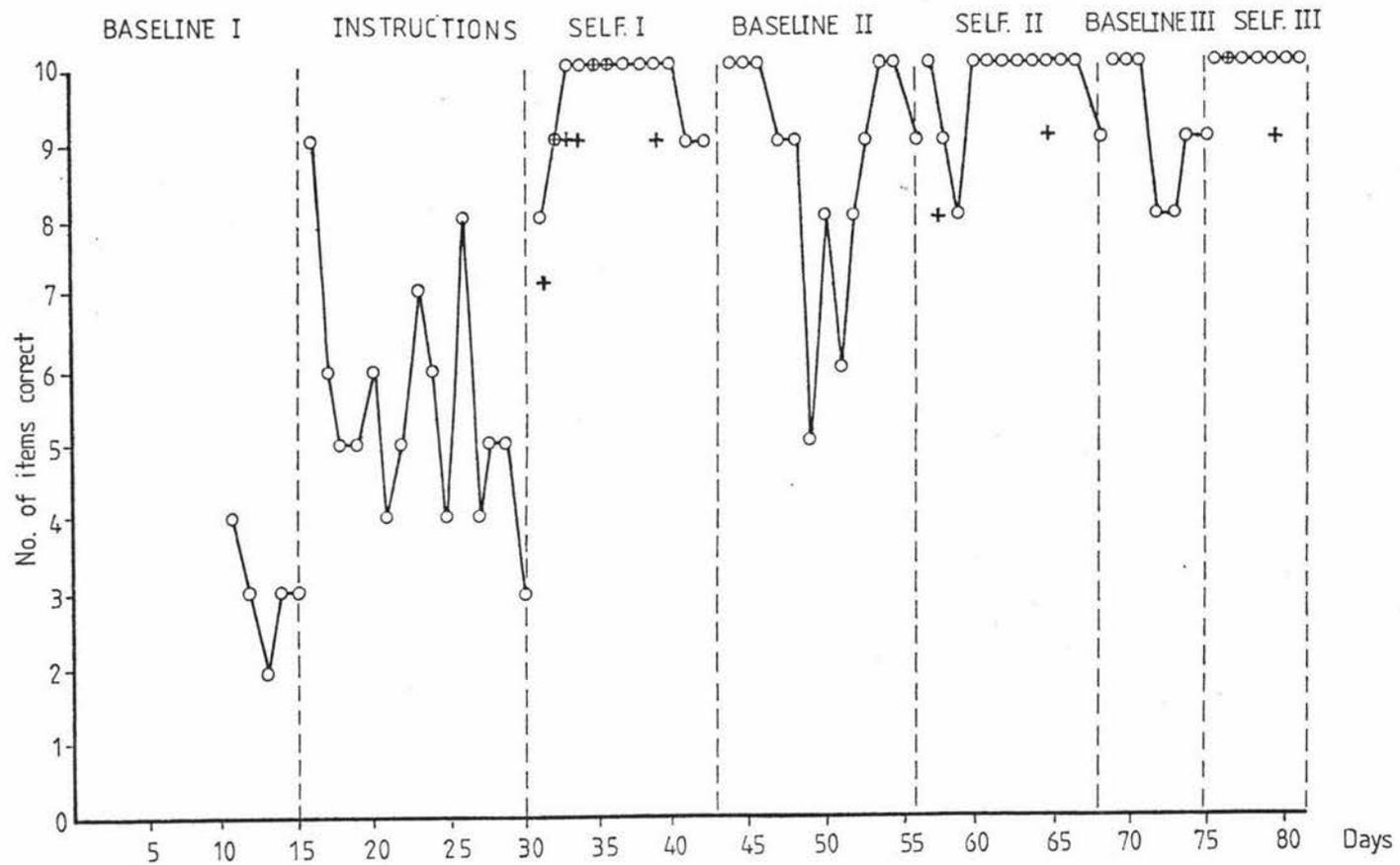


Fig.VIII. Subject 8, Sequence B

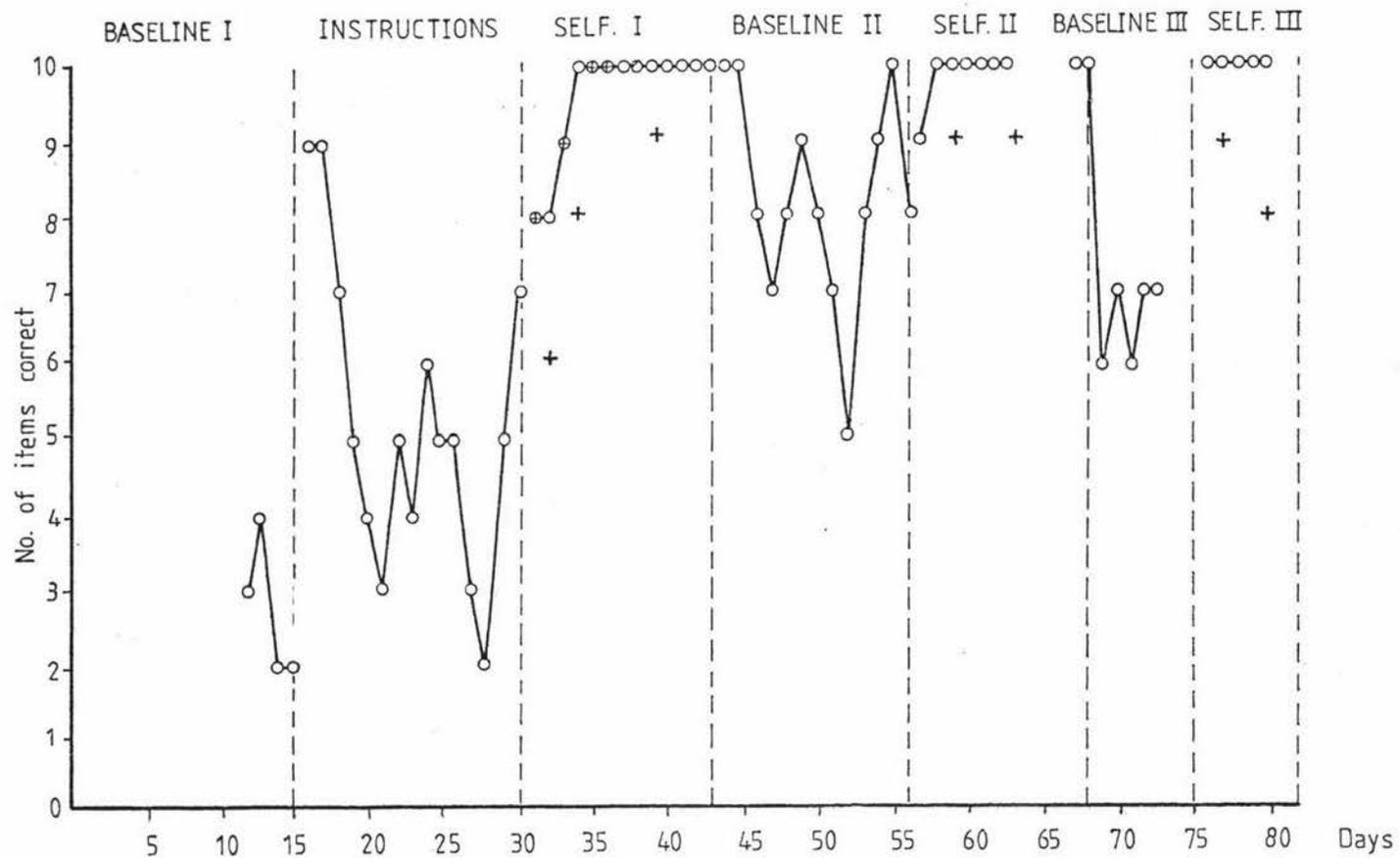


Fig.IX. Subject 9, Sequence B

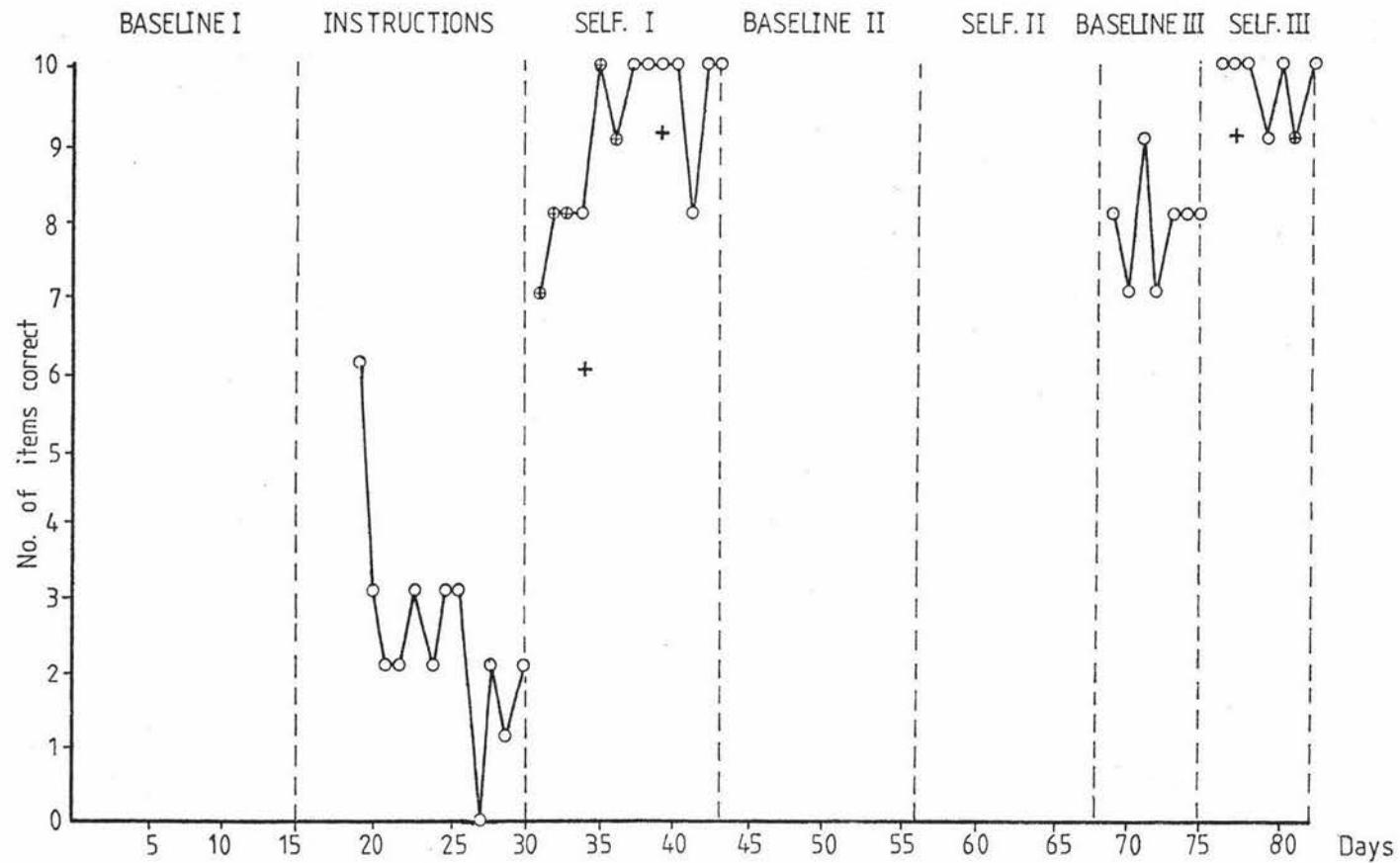


Fig.X. Subject 10, Sequence B

3.1.2 Combined Data

Figure 11 shows the mean daily performance for the Sequence A girls and Figure 12 the mean daily performance for the sequence B girls. Each data point was calculated using all individual girls' scores available for that day.

Figure 13 shows the mean performances but with each data point representing the average of every two days and with both Sequence A and Sequence B performances shown together. This was done to highlight trends within the data and to aid comparison between the two sequences.

3.2 ROOM CLEANING SCORES

In order to examine the major trends in the data, the combined graphs - figures 11 to 13 are considered first, followed by comments on individual variations.

The mean number of items correct during each experimental condition of Sequence A girls combined, and Sequence B girls combined are shown in Table 2.

Table 2

RESULTS SUMMARY

Mean number of items correct during
each experimental condition

SEQUENCE A GIRLS COMBINED		SEQUENCE B GIRLS COMBINED	
Baseline I	3.2	Baseline I	4.2
Instructions	3.4	Instructions	5.1
External Tokens	9.0	Self Evaluation I	9.5
Baseline II	7.5	Baseline II	8.0
Self Evaluation II	9.3	Self Evaluation II	9.8
Baseline III	6.1	Baseline III	7.8
Self Evaluation III	9.4	Self Evaluation III	9.8

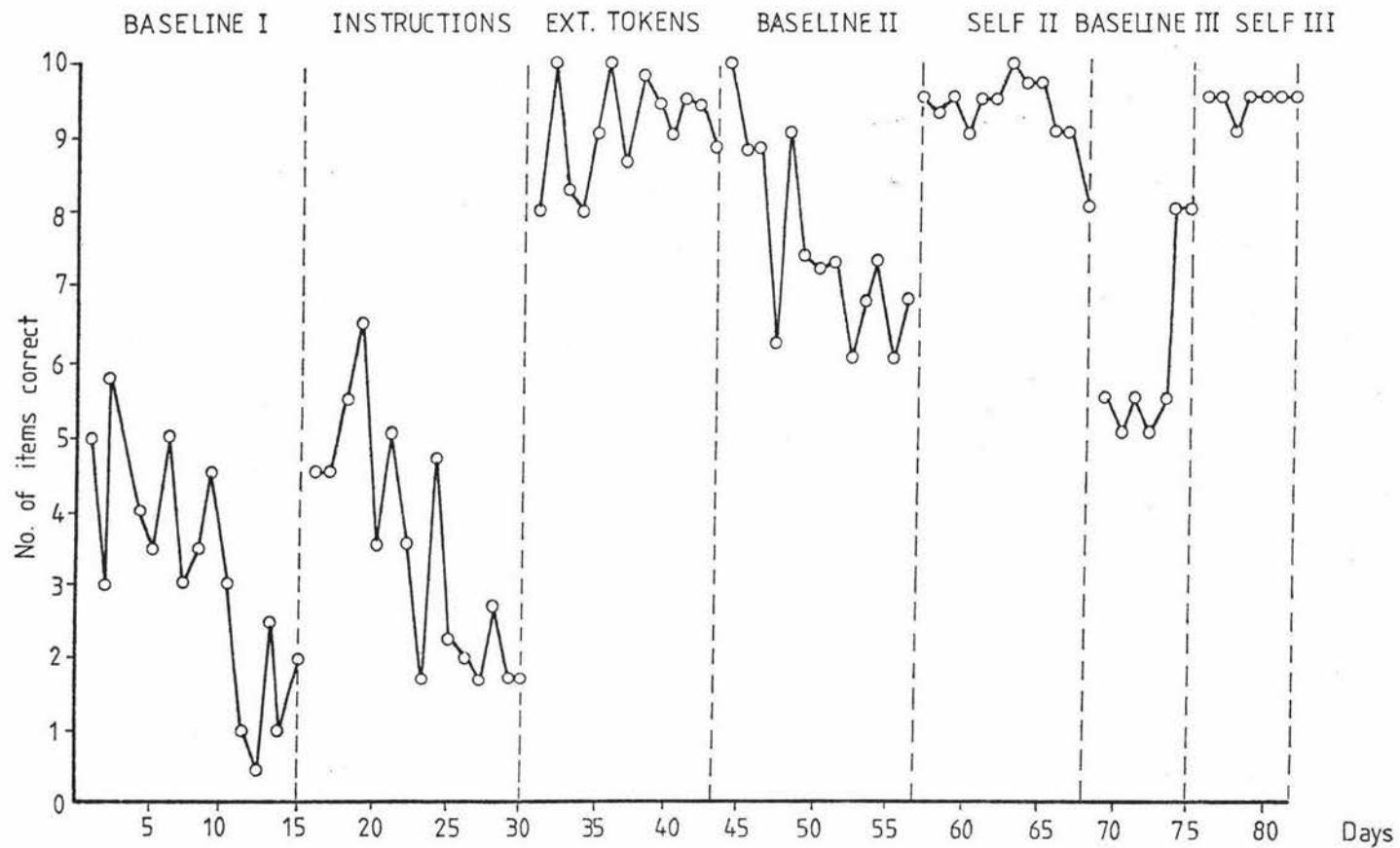


Fig.XI. Mean daily scores of Sequence A combined

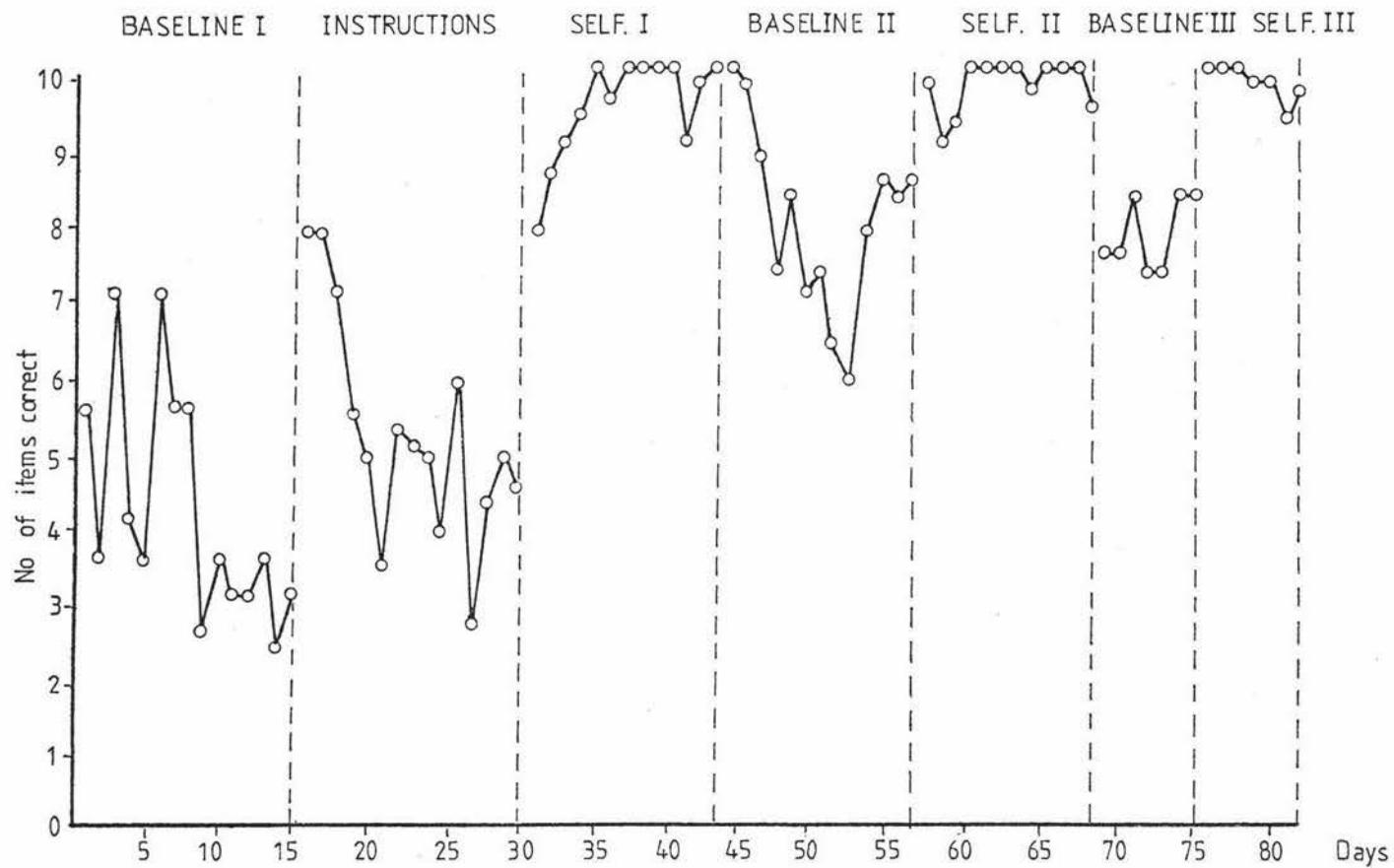


Fig. XII. Mean daily scores of Sequence B combined

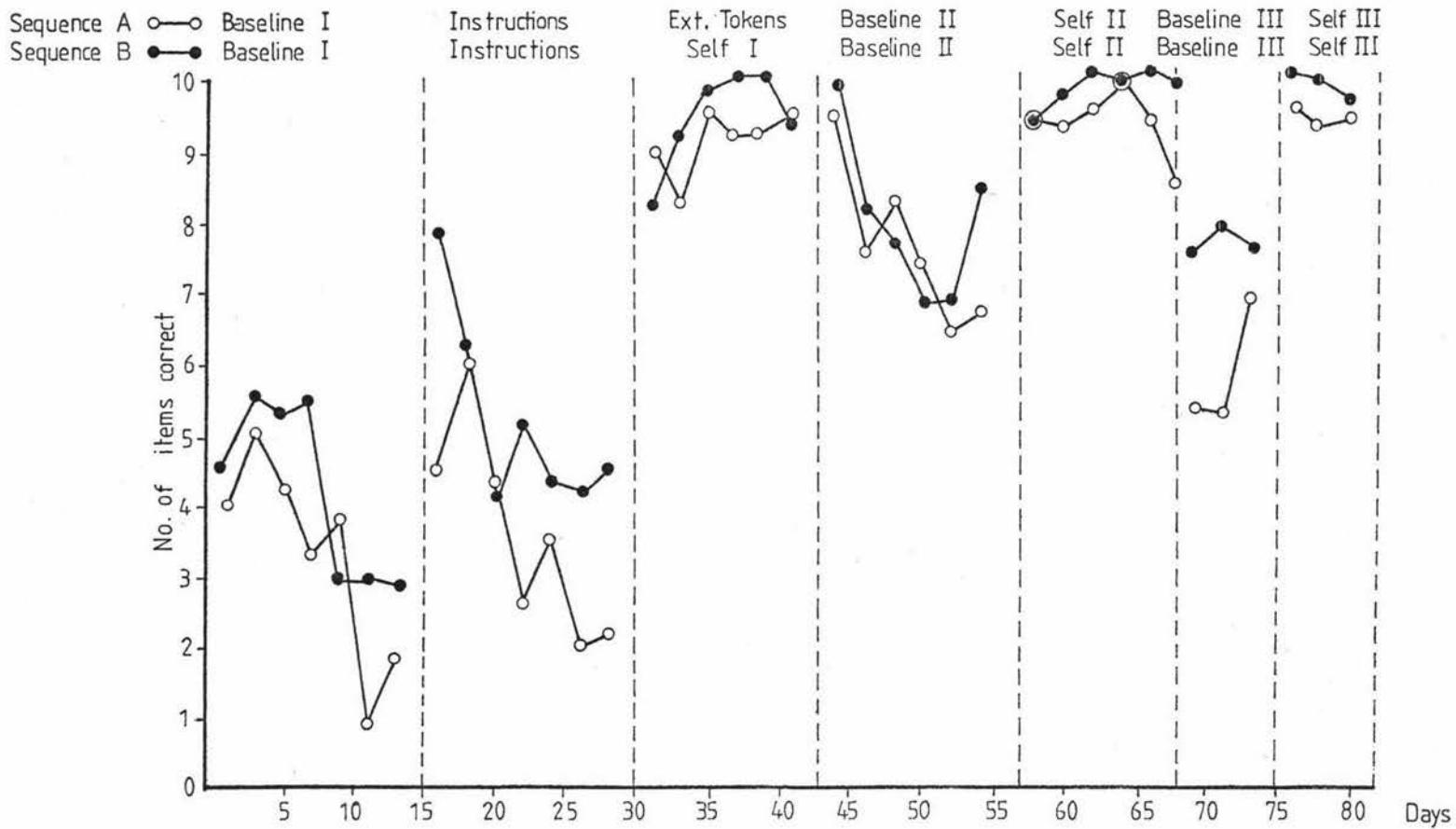


Fig. XIII. Mean daily scores of Sequence A combined & Sequence B combined

3.2.1 Baseline I (15 days)

During the first baseline phase there was a decreasing trend over time for both sequences. In all but one day the Sequence A scores were consistently lower than Sequence B. Sequence A scores began around 4.5 dropping to below 2.0 (average 3.2). Sequence B scores began about 5 dropping to about 3 (average 4.2). There were some wide variations in individual scores, with subject 2 scoring 1 four times and zero once during the last five days, and subject 6 scoring a possible ten on the third day.

3.2.2 Instructions (15 days)

When instructions were provided and demonstrations of the desired behaviour given, mean scores in both sequence groups increased substantially for the next few days, but declined with a few fluctuations over the following twelve days to near baseline levels. The initial reduction was greater with the Sequence B girls, but within a few days their score was similar to the Sequence A girls. The eventual decline was greater with Sequence A girls. Sequence A scores again dropped to about 2 (average 3.4) similar to the later Baseline I levels while the Sequence B scores levelled off at about 4.5 (average 5.1), consistently above the later Baseline I levels. There were some wide variations in individual scores, subject 2 had a peak score of 8 on day nine, and scored zero for the final two days. Subject 6 had wide variations from two scores of nine to two scores of two during the final five days, but had similar levels near end of Baseline I and Instructions. Subjects 6 and 7 shared a low trend throughout the

Instructions phase, and stayed at consistently higher levels than subject 2. Subject 9 was similar, although data on Baseline I was limited.

3.2.3 External Tokens/Self Evaluation I (13 days)

When the external token condition was initiated for Sequence A, room cleanliness scores rose rapidly and fluctuated daily between 8 and 10 (average for the whole phase 9.0). During this same period Sequence B began the Self Evaluation I condition and room cleanliness scores also rapidly increased, but to a higher level and with less fluctuations (average for the whole phase of 9.5). The Sequence A girls' scores tended to fluctuate more widely than those of Sequence B girls, except for subject 1, who usually scored consistently highly. Scores of subjects 6 and 10 in Sequence B were less consistent than the rest of this group.

3.2.4 Baseline II (13 days)

A similar decreasing trend was evident in the scores from both sequences, until the last three days. Both Sequence A and Sequence B scores decreased from around 9 to around 6 over the first nine days. In the last four days Sequence A scores continued to drop slightly to around 6.5 (average for the whole phase of 7.5), while Sequence B scores rose to around 8.5 (average for the whole phase of 8.0). Both sequences A and B remained consistently higher than their respective Baseline I levels.

Individual scores for Sequence A girls showed that the group effect was caused mostly by subjects 2 and 4,

although some drop was shown by all individuals. Individual scores in Sequence B show that subject 7, and in the first 8 days, subject 6, had the most marked effect on reducing the scores, although all subjects responded overall lower than during the Self Evaluation I phase.

3.2.5 Self Evaluation II (12 days)

When the Self Evaluation II phase commenced, both groups rapidly gave improved scores. Sequence A mean scores were always 9 or above except for the last day, (average 9.3), and Sequence B mean scores were also always 9 or above but were overall, higher (average 9.8). After the third day, Sequence B was consistently higher than Sequence A. These levels were higher overall for each sequence than those obtained in their previous treatment conditions, (external tokens for Sequence A and Self Evaluation I for Sequence B). Individual scores show that for Sequence A again the scores of subject 2 had the most marked effect on lowering the scores. Subject 4 did have a marked effect on two days. Individual scores of all Sequence B girls showed an ascending trend over the first three days, (except for subject 6 who scored 10 on the first day) and then levelled out at ten items correct per day.

3.2.6 Baseline III (7 days)

During the third baseline there was a marked decline in the daily mean score for all subjects. However, after the initial drop on the first day, the scores for Sequence A girls stayed just above 5.0 for the first five days and

then rose to around 8.0 for the last two days (averaging 6.1 over the phase), whereas the scores for the Sequence B girls stayed at about 8.0 throughout (average 7.8). This resulted in the girls from both sequences levelling off close to that in the latter stage of their respective Baseline II levels, again at a lot higher level than during Baseline I. Individual scores again had a marked effect. At this stage only two girls in the Sequence A group were still in the home. Subject 2 lowered the average and produced the marked upswing in the latter part of the phase. Subject 5 maintained higher scores, similar to the girls in Sequence B. Apart from subject 6, the Sequence B girls tended to maintain scores at a constant level throughout the phase, but the level for each girl varied from about 6.5 (subject 9) to about 9.0 (subject 8). Baseline III was discontinued after 7 days, due to a very strong negative reaction to the procedure from the girls.

3.2.7 Self Evaluation III (7 days)

This final self evaluation condition was initiated during the last week of term after the study had originally been scheduled for conclusion. Each of the two Sequence A girls scores returned very closely to their Self Evaluation II levels, (averaging a score of 9.4). The mean score of the Sequence B girls returned to high Self Evaluation II levels although there was a slight trend down in the last few days (overall average 9.8). This was caused largely by subject 6 dropping her score, and by high scoring subject 9 leaving the home.

3.3 ACCURACY OF SELF EVALUATION

Table 3 shows the mean and range figures expressed as percentage accuracy of Self Evaluations for Sequence A girls combined, and Sequence B girls combined. The percentage accuracy of Self Evaluations was calculated in a similar manner to that used for the interobserver reliability checks. The number of agreements item by item between a girl's self evaluation and a residential Social Worker's evaluation was divided by the total number of agreements plus disagreements and multiplied by 100.

3.3.1 Self Evaluation I

During the first seven days of this condition, points for accuracy of matching were given daily to Sequence B girls. Accuracy points were given on the basis of a $\frac{1}{2}$ point for each item of the girl's self evaluation that matched the residential Social Worker's evaluation of her. An 80 percent accuracy criterion (the same as used by Wood and Flynn, 1978) was either met, or exceeded in all cases. No return to daily matching of evaluations, or use of accuracy points was needed later in this phase. Spot checking revealed that the 80 percent criterion levels were maintained by all girls following their initial training in accurate reporting of self evaluations. The cross checkpoint at day 9 of Self Evaluation I on each Sequence B girl's individual graph indicates the percentage level of accuracy maintained.

Table 3

ACCURACY OF SELF EVALUATIONS SUMMARY

(Mean and range figures expressed as percentage accuracy)

	MEAN	RANGE	NO. REACHING CRITERIA
SEQUENCE A GIRLS COMBINED:			
Self Evaluation II (random checks)	83.8	50-100	2 out of 4
Self Evaluation III (random checks)	83.0	70-100	1 out of 2
SEQUENCE B GIRLS COMBINED:			
Self Evaluation I (daily matching)	94.6	80-100	5 out of 5
Self Evaluation II (random checks)	88.8	80- 90	4 out of 4
Self Evaluation III (random checks)	92.5	80-100	4 out of 4

3.3.2 Self Evaluation II

Checking for accuracy in this condition was done on a random basis by the staff. The random checks consisted of a staff member selecting one girl from each bedroom per day and checking her self evaluation item by item with the staff member's evaluation. The girl was then informed of how many items in her evaluation matched that of the staff member. The inaccuracies were pointed out to the girl and verbal encouragement was given for accurate reporting. There were no points given for accuracy, or deducted for inaccuracy. Points were given on the basis of the girl's

evaluation of her room cleaning. The results of these random checks are presented in Table 3. These show that all the girls in Sequence B who had earlier been trained in, and given points for accurate matching, still continued to meet the criterion of 80 percent. The Sequence A girls who had the External Token system previously, with no training in accurate matching failed to meet an accuracy level of 80 percent in 2 out of 4 subjects.

3.3.3 Self Evaluation III

Checking for accuracy in this condition was done on a random basis, identical to that used in Self Evaluation II. This condition was instigated because during Baseline III when the opportunity for continuing self evaluation had been removed there had been considerable negative reaction from the girls. The previously invested responsibility for self evaluation and the trust implicit in it, being removed after it had become established, was cited by the girls as the basis for their negative reaction to the return to baseline procedures. When Self Evaluation III was commenced girls were told that if they were to continue with self evaluation they would need to be more accurate than they had been during Self Evaluation II. The random check results of Self Evaluation III, shown in Table 3, indicate an overall 4 percent improvement in accuracy compared with Self Evaluation II for Sequence B girls with all maintaining at least 80 percent accuracy. There was nearly a 1 percent decrement in accuracy compared with Self Evaluation II for Sequence A girls with only 1 out of the 2 girls maintaining at least 80 percent accuracy.

CHAPTER 4

DISCUSSION

4.1 The goal of this study was to develop an accurate behavioural self evaluation and reporting procedure which established and maintained appropriate levels of behaviour with the target group of delinquent girls in residential care. Room cleaning behaviour was the dependent variable and a series of procedures similar to those used by Wood and Flynn (1978) was adopted.

4.2 SUMMARY OF STUDY RESULTS

4.2.1 An appropriate level of room cleaning behaviour was achieved by the use of external tokens as contingent reinforcement. Subjects in Sequence A received this treatment and averaged 9.0 items correct during a 13 day period, with individual fluctuations between 8 and 10 items correct. During a subsequent baseline period, the mean score dropped to 7.5 items correct, but maintained a four point improvement over a previous baseline period.

4.2.2 An appropriate level of room cleaning behaviour was achieved by training girls to evaluate their own behaviour and requiring them to accurately match the evaluations of an observer. Points were given contingent on accuracy and room cleanliness. Subjects in Sequence B received this treatment and averaged 9.5 items correct per day during a 13 day period. There were less fluctuations in scores with this group, than with the group who received external tokens.

During a subsequent baseline period the mean score decreased to 8.0 items correct but maintained a four point improvement over a previous baseline period.

4.2.3 An appropriate level of room cleaning behaviour was achieved by having both Sequence A and Sequence B girls carry out self evaluations of room cleanliness, with matching of scores being done on a random basis and contingent reinforcement being self-determined. Mean scores for the two sequences during this self evaluation period were 9.3 for Sequence A and 9.8 for Sequence B. The girls in Sequence A who had not received points contingent on accurate matching previously, tended to be less accurate than those in Sequence B who had.

4.2.4 An appropriate level of room cleaning was achieved by both Sequence A and Sequence B girls carrying out self evaluations of room cleanliness with matching of scores being done on a random basis and no contingent points being available. Mean scores during this self evaluation period were 9.4 for Sequence A and 9.8 for Sequence B.

4.2.5 Throughout the study the low level recorded at Baseline I was never returned to during either Baseline II or Baseline III. The mean of Baseline II for both groups was similar, and the mean for Baseline III being higher for Sequence B than Sequence A.

4.2.6 The use of instructions alone in the second phase of the study initially increased room cleaning behaviour to appropriate levels but it returned to near baseline levels when reinforcement contingencies were not given. One factor

to emerge from this phase was that instructions appeared to have a transient effect on behaviour.

4.2.7 The most important factor to consider in this summary of results is the variation in individual scores. Mean scores have tended to give a rather distorted picture of the behaviour during the study. For instance Subject 2 consistently fluctuated in scores throughout most phases. In Baseline I she was one of only two subjects in her sequence completing the phase, and she scored 1 four times and zero once. Comparatively, subject 6 scored a possible 10 during this phase. During the course of the external tokens phase and the three self evaluation phases, subject 2 fluctuated between 7 and 10 points per day consistently. Subjects 4 and 5 started at 7 points per day and gradually increased to fluctuate between 9 and 10 points per day. Subjects 6 and 10 ranged between 8 and 10 points per day. Subjects 7, 9, 9 and 1 rose to 10 points per day and stayed there consistently. Individual variation thus needs to be borne in mind when interpreting mean scores.

4.2.8 Although appropriate levels of room cleaning behaviour were maintained during the treatment conditions the other aspect of the study concerned accurate behavioural self evaluation.

Sequence B girls received points for accuracy of matching the observers score during self evaluation I. This procedure proved very effective in ensuring that accuracy was achieved. The points for accuracy condition did not remain in force longer than a week as the 80% criterion of accuracy was either met or exceeded in all

cases. In the subsequent Self Evaluation II and Self Evaluation III conditions when only random checks on accuracy were made, these girls still continued to keep to the 80 percent criterion.

Sequence B girls never had specific training in accuracy of matching with contingent points. As a result half of them never reached an 80 percent level of accuracy. Despite the lack of training for this group, one girl, subject 5 averaged a personal accuracy level of 97.5 percent. Once again an illustration that mean scores, especially for small groups, should be regarded with caution. Looking at other individuals, subject 2 who had no accuracy training, never averaged above 70 percent accuracy, and subject 4 only averaged 65 percent during the course of two random spot checks.

4.2.9 In summary of these results, two conclusions can be drawn:

- 1) Sequence A girls, did not have accuracy training, but instead passively received the judgments of an observer at the external tokens phase. They reached criterion levels of accuracy on spot checks in only 50% of cases, hence their self evaluated scores can be viewed with a certain degree of doubt. It is highly likely that during Self Evaluation II and III these girls were inaccurately scoring themselves.
- 2) Sequence B girls, did have accuracy training, and all reached or exceeded criterion levels during

it, and throughout the random checks that followed. Their scores could be considered as accurate representations of the effect of the experimental conditions. In other words one could view with some optimism that in a residential programme both accurate behavioural self evaluation, and appropriate levels of a target behaviour can be achieved by using a matching procedure with points contingent on accuracy, and points for completion of items specified for the target behaviour.

4.3 COMPARISON WITH WOOD AND FLYNN RESULTS

i) There were no major variations among the youths in each sequence group throughout the experimental conditions. This is in marked contrast to this systematic replication.

ii) Wood and Flynn had a lower Baseline I level than this study. The higher baseline I level noted in this study could have been due to the residual effect of the token programme still operating for dormitory behaviour.

iii) The pattern observed during Instructions phase was similar for both studies.

iv) A similar ascending pattern of data was obtained in both studies during the external tokens and Self Evaluation I phases.

v) Return to baseline after external tokens/self evaluation I phase was quite different between the studies.

Wood and Flynn found that subjects who had received External tokens had lower subsequent baseline levels than youths in the concurrent self evaluation I phase, and used this result to suggest that teaching self evaluation within a token system was more effective than an externally imposed token system alone. In this study there was no similar effect observed.

vi) A similar rise in scores at Self Evaluation II was observed in both studies.

vii) Baseline III for Wood and Flynn remained steady and at a high level. In this study it returned to previous Baseline II levels and in some cases marginally below them.

viii) The major difference was in the time taken for the phases. It is quite unlikely that Baseline periods in this study of 13 and 7 days respectively represent any adequate test of resistance to extinction. Wood and Flynn had 60 days Baseline II for Sequence B youths, and 22 days Baseline III for all youths. These long Baseline periods represent an adequately long test of resistance to extinction on performance standards.

ix) Wood and Flynn used accuracy points for matching with both sequences, they reported a 93% accuracy during the accuracy points phases, and in following spot checks the 80 percent criteria was always met. In comparison, this study had similar accuracy levels when Sequence B received points for accurate matching. However, as I have mentioned the randomly checked Sequence A who never had such training could well have been constantly outside criterion levels for most of the time.

4.4 RECORDING AND RELIABILITY

4.4.1 Target Behaviours

The dependent variable for both studies was room cleaning behaviour. The criteria used at the Living and Learning centre were extremely circumscribed, with fifteen definitions being modified from Fixsen et al., (1972). The criteria used in the replication were derived from Fixsen et al., (1972) and Wood and Flynn (1977) with considerable influence from Weir and Ford (1977). The development of the ten point criteria was adopted at the Girls' Home with an emphasis on total cleanliness of a room which included wall, floors, windows, curtains, clothing and all furniture. In general it was broader in scope, but nevertheless finally evolved as being unambiguous. The Mark III version of the criteria used is included in the text (Table 1), and the two previous versions worked on during the pilot study undertaken in the final two weeks of the second term, are included in the appendix. The scoring sheet used for recording items correct or incorrect, and for recording inter-observer reliability checks was similarly finalised during the pilot study and is included in the appendix.

4.4.2 Observations

The observations at the Living and Learning centre were made every day by an adult observer while the youths were at school. During Baselines, the youths were not told at any time that their rooms were checked. The graphs produced are all plotted points of an adult observer. The observations at the Girls' Home were made by staff each day during Baselines, Instructions, and External Tokens phases.

The girls' self evaluations are plotted during the three self evaluation phases, and the matching and random checks made by staff are illustrated by crosses on graphs. During Baselines at the Girls' Home some unavoidable awareness that rooms were being checked occurred, as there were always some girls resident in the home during the day.

Interobserver reliability checks proved one of the most difficult aspects of the replication. The calculation system used was similar to that used by Wood and Flynn (1978). They reported an average reliability of 90 percent over the nine checks they made, while in this study an average reliability of 87.4 percent was obtained across the fourteen reliability checks made. The necessity to rigorously adhere to criteria when making checks was established during the pilot study at the Girls' Home, when training in reliability checking was done with all the staff. In addition to the 14 staff/staff reliability checks done during the study, there were seven days of matching reliability checks done by staff and girls during Self Evaluation I, and a further two days of random checks done by staff on girls during each of the Self Evaluation II and III phases. The results of the reliability checks are included in the appendix for staff/staff, and the results of reliability checks for staff/girl are also included in the appendix. The mean accuracy figures for girls are in Table 3.

4.5 PROCEDURES

4.5.1 Baselines I, II and III

In all these phases of the Wood and Flynn study there

were no scheduled consequences for room cleaning behaviour, but during these periods, the residential token economy was maintained and reinforcement, including recreation, was available for other behaviours. It is difficult to assess unequivocally, that there were no residual effects of this reinforcer maintaining behaviour during Baseline II and III periods since recreation and access to it was initiated by the study.

A residual effect of the residential token economy in this study, especially during Baseline I is accepted. The subjects were predominantly new admissions who were receiving three points per day (morning, afternoon and evening) for general dormitory behaviour. It has been noted that this residual effect decreased over the period of Baseline I to final levels of between 1 and 3 items correct per day. It remains unlikely that this residual effect had any bearing on the Baseline II results, and definitely not in Baseline III as there were no points of any kind available.

4.5.2 Self Evaluation

The Self Evaluation I procedure followed in this study was identical to Wood and Flynn who used the matching and fading procedure of Drabman et al., (1973). The major variation was in the allocation of points. Wood and Flynn had an allocation of up to 15 points for degree of room cleanliness, and up to 15 points for accuracy of evaluation. This meant that during this sequence youths could earn up to 30 points, double the amount youths in the concurrently

running External Token phase. This reinforcement condition was considered inequitable for this study, as it appeared to give an advantage to one sequence over another for access to the reinforcement. In this study a half point was allocated for degree of room cleanliness, and a half point for accuracy of evaluation. In this way access to reinforcement was not weighted in favour of either sequence. In actual fact it meant that a girl in Self Evaluation I had to perform two actions to earn 1 point - namely accuracy and completing room cleaning items, compared with External Tokens girls who performed room cleaning alone. This factor was commented on by the girls at the time. They realised there was a different contingency for self evaluators, but since this condition only remained in force for a week it had no long lasting effect, if anything, self evaluating was considered a high status activity.

In the Self Evaluation II and III conditions of this study girls from Sequence A were not trained in accuracy of matching, in order to give a comparison with Sequence B girls, and to attempt a replication of Santagrossi et al., (1973). By use of random checks, it was possible to investigate the effectiveness of both an accuracy training procedure and a straight self evaluation procedure.

4.5.3 Reinforcement Considerations

The Wood and Flynn study used access to recreation as the generalized reinforcer. This access was moderated by youths averaging, or earning a specified number of points each day. It is unclear if the recreation points earned in

experiment were the only way recreation could be earned, or a bonus to recreation points which were available for other behaviours. During Baselines, the residential token economy was maintained and reinforcement, including recreation, was available for other behaviours. During the external token system the youth had to earn a specified number of room cleaning points to attend a recreation period, but these points were completely separate from the usual points used to operate the residence, and were called 'recreation points'. Other points received during the day for behaviours other than room cleaning did not count towards recreation. Thus recreation during this phase was contingent on room cleaning alone. During self evaluation the youths had to earn both, points for degree of room cleanliness and points for accuracy of evaluation. There is no indication that access to recreation was only contingent on room cleaning, or available for other behaviours. This degree of ambiguity creates difficulty in assessment of any residual effect of the residential token economy maintaining room cleaning behaviour.

This study used money as the primary reinforcer for room cleaning behaviour. The money earned was a bonus to the normal pocket money earnings from the residential token economy. The secondary reinforcer was increased access to town outings and weekends home created by higher earnings. The amount of financial reinforcement was increased from 3 cents per day at Baseline I, to 10 cents per day during the external tokens and self evaluation I and II. During Baseline III, and Self Evaluation III this

financial reinforcement was completely withdrawn. The residual effect of the residential token economy which appeared during the initial stages of Baseline I diminished over time, and in the final two phases of Baseline III and Self Evaluation III no reinforcement was available to effect room cleaning behaviour.

4.6 SUMMARY AND CONCLUSIONS

4.6.1 Currently there are a number of critical questions being asked about treatment programmes, particularly residential placements for offenders. The issues of social validity, methodological approaches taken, and the long term effect of all these is being continually reviewed. This tends to force the practitioner to search for methods of improvement to residential programmes.

4.6.2 Token economies have enjoyed wide success as a treatment technique for institutionalised persons, but there is increasing evidence that such programmes, when withdrawn result in a deterioration in the maintenance of behaviour change. A development has been to incorporate self evaluation procedures into token programmes.

4.6.3 Effective means of maintaining appropriate behaviour by use of 'accuracy matching' have been reported by Drabman et al., (1973). Wood and Flynn (1978) used such a procedure with delinquent youth in a residential token economy, where both accurate self evaluation and appropriate room cleaning behaviour were achieved and maintained. Although the youths were resident for a long

time, good treatment effects were observed. However, this time scale suggests that a self evaluation phase would need to be included much earlier to have wider applicability as many residential token programmes do not have the luxury of such a time scale.

4.6.4 In this systematic replication a time constraint of three months from initiation to conclusion was paramount. Thus within this framework a tentative confirmation of Wood and Flynn's Sequence B results has suggested itself. In addition, the comparison between accurate self reporting by Sequence B girls and often inaccurate reporting by Sequence A girls, lends further support to the work of Santagrossi et al., (1973).

4.6.5 This study was an initial step in trying to develop within the Residential Programme of a Department of Social Welfare institution an accurate behavioural self evaluation and reporting procedure which established and maintained appropriate levels of behaviour which could be an aid to transfer of treatment gains into the open community setting.

APPENDICES

APPENDIX A

ROOM CLEANING CRITERIA MK I

1. Bed - made up with - 1 under blanket, 2 sheets, 1 or 2 top blankets, 1 bedspread, 1 quilt (if available).
2. Sheets and blankets tucked in all the way around the bed. Top sheet folded over 30 cm from head of bed. Pillow in centre of bed at head, edges at least 15 cm from sides of bed. Bedspread tucked in at foot of bed, sides hanging down to floor level, folded over pillow with fold line visible under pillow. Quilt placed on top of bed not overhanging the sides.
3. Wardrobe clothing - All non-folded clothes, to be hung on hangers, all hangers to be on rail.
4. Wardrobe footwear - Shoes, toe sandals, etc, to be placed in pairs, sides touching, toes touching rear wall of wardrobe.
5. Bedside Cabinets, Sets of drawers (Interior) - all clean clothes folded, other objects sorted in an orderly fashion.
6. Bedside Cabinets, Sets of drawers (Exterior) - tops to be dusted, no objects within 10cm of any edge, cabinets and drawers no further than 10cm from wall.
7. Posters, pictures etc., - to be pinned to wall hangings only. Nothing to be sellotaped to walls and wallpaper.
8. Curtains to be hanging from tracks attached to runners, drawn back during day to within 60cm of window frame. Window sills and surrounds to be free of dust and dirt. Windows and mirrors to be free of dust and dirt.
9. Floors to be free of dust and dirt. Walls to be free of dirt and cobwebs. Ceilings to be free of dirt and cobwebs. Mirror to be free of dirt.
10. Towels and face cloths to be hung on rails by beds, ends to be even. Dirty linen bags to be hung on hooks inside wardrobes. No clothing or personal objects on beds, floors or furniture, apart from personal objects on bedside cabinets and sets of drawers.

ROOM CLEANING CRITERIA MK II

1. Bed - made up with - 1 under blanket, 2 sheets, 1 or 2 top blankets, 1 bedspread, 1 quilt (if available). Sheets and blankets tucked in all the way around the bed, top sheet folded over 30 cm from bed head.
2. Bed - Pillow in centre of bed at head, edges at least 15cm from sides of bed. Bedspread tucked in at foot of bed, sides hanging down to floor level, folded over pillow with fold line visible under pillow. Quilt placed on top of bed not overhanging the sides.
3. Wardrobe Clothing - All non-folded clothes to be hung on hangers, all hangers to be on rail.
4. Wardrobe Footwear - shoes, toe sandals etc, to be placed in pairs, sides touching, toes touching rear wall of wardrobe.
5. Bedside Cabinets, Sets of Drawers (Interior) - All clean clothes folded, other objects sorted in an orderly fashion.
6. Bedside Cabinets, Sets of Drawers (Exterior) - Tops to be dusted, no objects within 3cm of any edge, cabinets and drawers no further than 5cm from any wall.
7. Posters, Pictures - To be pinned onto either the wall hanging, or the display boards. Nothing to be sellotaped to walls or wallpaper.
8. Curtains - to be hanging from tracks, attached to runners drawn back during day to within 60cm of window frame. Window sills and frames to be free of dust and dirt. Windows to be free of dust and dirt.
9. Floors - to be free of any dust and dirt. Mirrors to be free of dust and dirt.
10. Towels - Towels and face cloths to be hung on rails by beds, ends to be even. Dirty linen bags to be hung on hooks inside wardrobes. No clothing, or personal objects on bedside cabinets and sets of drawers.

PALMERSTON NORTH GIRLS HOME

SELF MONITORING DEVELOPMENT PROGRAMME

INTER-OBSERVER RELIABILITY CHECKS

GIRLS NAME..... SCORERS NAME.....

DATE

PHASE -

Baseline I	
Instructions	
Ext. Tokens	
Self Eval. I	
Baseline II	
Self Eval. II	
Baseline III	

BEHAVIOUR CATEGORY

1. ROOM CLEANING BEHAVIOUR

CRITERIA (if met a ✓ in box, if not met a 0 in box.)

CRITERIA	comments	I/OSAG/OC	AG/NO	Disag.
1) BED - Pillow/Bedsread/Quilt				
2) BED - Blankets/Sheets				
3) WARDROBE - Clothing				
4) WARDROBE - Footwear				
5) BEDSIDE CABINETS/DRAWERS- Int.				
6) BEDSIDE CABINETS/DRAWERS- Ext.				
7) POSTERS/WALLHANGINGS				
8) CURTAINS/WINDOWS/FRAMES				
9) FLOORS/ MIRRORS				
10) LAUNDRY BAGS				
SCORE				
PERCENTAGE CORRECT				

APPENDIX C

INTEROBSERVER RELIABILITY CHECKS

The mean percentage interobserver reliability for each check done in each phase is presented:

CONDITION	CHECK	MEAN	PHASE MEAN
BASELINE I	1	72.0	
	2	85.0	
	3	85.7	
	4	91.3	
	5	95.5	
	6	80.0	84.9
INSTRUCTIONS	1	72.5	
	2	85.0	78.8
EXTERNAL TOKENS/ SELF EVALUATION I	1	91.3	
	2	95.8	93.5
BASELINE II	1	86.5	86.5
SELF EVALUATION II	1	90.5	90.5
BASELINE III	1	88.3	88.3
SELF EVALUATION III	1	89.3	89.3

Mean of all Phases = 87.4

Range of all Phases = 78.8 - 93.5

APPENDIX D

ACCURACY OF SELF EVALUATIONS

RELIABILITY CHECKS BY STAFF OF GIRLS

(Figures expressed as percentage accuracy)

SEQUENCE A GIRLS

SELF EVALUATION PHASE	SUBJECT	RANDOM CHECKS - NO POINTS FOR ACCURACY OF MATCHING		
		CHECK 1	CHECK 2	MEAN
II	1	100	100	100*
	2	60	80	70
	4	50	80	65
	5	100	100	100*
				83.8
III	2	70	70	70
	5	90	100	95*
				83.0

* reached criterion levels

ACCURACY OF SELF EVALUATIONS

 RELIABILITY CHECKS BY STAFF OF GIRLS
 (figure expressed as percentage accuracy)
SEQUENCE B GIRLS

SELF EVALUATION I Matching and points for accuracy of matching

SUBJECT	CHECKS							MEAN
	1	2	3	4	5	6	7	
6	80	80	100	80	100	100	100	91.4
7	100	100	90	100	100	100	100	98.6
8	90	100	90	90	100	100	90	94.3
9	100	80	100	80	100	100	90	92.9
10	100	100	100	80	100	100	90	95.7
								94.6

SELF EVALUATION II Random checks - No points for accuracy of matching

SUBJECT	CHECK		MEAN
	1	2	
6	90	90	90
7	90	80	85
8	90	90	90
9	90	90	90
			88.8

SELF EVALUATION III Random checks - No points for accuracy of matching

SUBJECT	CHECK		MEAN
	1	2	
6	90	100	95
7	100	90	95
9	90	80	85
10	90	100	95
			92.5

All of these subjects maintained criterion level of 80% accuracy or better.

REFERENCES

- Allison, T.S., Kendall, S., & Sloane, D. New directions in a juvenile hall setting. In J.S. Stumphauzer (Ed.). Progress in behavior therapy with delinquents. Springfield, IL: Charles C. Thomas, 1979.
- Ayllon, T., & Azrin, N. The token economy. New York: Appleton-Century-Crofts, 1968.
- Ayllon, R., & Milan, M.A. Correctional rehabilitation and management: A psychological approach. New York: Wiley, 1979.
- Baer, D.M. In the beginning, there was the response. In E. Ramp & G. Semb (Eds.), Behaviour analysis: Areas of research and application. Englewood Cliffs, N.J: Prentice-Hall, 1975.
- Baer, D.M., Wolf, M.M., & Risley, T.R. Some current dimensions of applied behavior analysis. Journal of Applied Behavior Analysis, 1968, 1, 91-97.
- Bandura, A., & Perloff, B. Relative efficacy of self-monitored and externally imposed reinforcement systems. Journal of Personality and Social Psychology, 1967, 7, 111-116.
- Bassett, J.E., Blanchard, E.B., & Koshland, E. Applied behavior analysis in a penal setting: Targeting "free world" behaviors. Behavior Therapy, 1975, 6, 639-648.
- Beck, S., & DeVoge, T. Generalization of treatment effects for Juvenile Delinquent treatment settings. Journal of Juvenile and Family Courts, Feb. 1978, 9-12.
- Bolstad, O.D., & Johnson, S.M. Self-regulation in the modification of disruptive classroom behavior. Journal of Applied Behavior Analysis, 1972, 5, 443-454.
- Braukmann, C.J. & Fixsen, D.L. Behavior modification with delinquents. In M. Hersen, R.M. Fisher, & P.M. Miller (Eds.), Progress in behavior therapy. New York: Academic, 1975.
- Braukmann, C.J., Fixsen, D.L., Kirigin, K.A., Phillips, E.A., Phillips, E.L., & Wolf, M.M. Achievement Place; The training and certification of teaching parents. In W.S. Wood (Ed.), Issues in evaluating behaviour modification. Champaign, IL: Research Press, 1975.
- Braukmann, C.J., Fixsen, D.L., Phillips, E.L., & Wolf, M.M. Behavioral approaches to treatment in the crime and delinquency field. Criminology, 1975, 13, 299-311.

- Braukmann, C.J., Kirigin, K.A. & Wolf, M.M. Achievement Place: The researcher's perspective. Paper presented at the meeting of the American Psychological Association, Washington D.C., September 1976.
- Burchard, J.D., & Harig, R.T. Behavior modification and juvenile delinquency. In J. Leitenberg (Ed), Handbook of behavior modification and behavior therapy. Englewood Cliffs, N.J: Prentice-Hall, 1976.
- Davidson, W.S., & Seidman, E. Studies of behavior modification and juvenile delinquency: A review, methodological critique, and social perspective. Psychological Bulletin, 1974, 81, 998-1011.
- Davidson, W.S., and Wolfred, T.R. Evaluation of a community-based behavior modification programme for prevention of delinquency: The failure of success. Community Mental Health Journal, 1977, 13(4), 296-306.
- Drabman, R.S., Spitalnik, R., & O'Leary, D.K. Teaching self control to disruptive children. Journal of Abnormal Psychology, 1973, 82, 10-16.
- Emery, R.E., & Marholin, D. An applied behavior analysis of delinquency. American Psychologist, 1977, 32, 860-873.
- Fixsen, D.L., Phillips, E.L., and Wolf, M.M. Achievement Place: The reliability of self reporting and peer-reporting and their effects on behavior. Journal of Applied Behavior Analysis, 1972, 5, 19-30.
- Fixsen, D.L., Phillips, E.L., and Wolf, M.M. Mission-oriented behavior research: The teaching family model. In A.C. Catania & T.A. Brigham (Eds.), Handbook of applied behavior analysis: Social and instructional processes. New York: Irvington, 1978.
- Frederiksen, L.W., and Frederiksen, C.B. Teacher-determined and self-determined token reinforcement in a special education classroom. Behavior Therapy, 1975, 6, 310-314.
- Hundert, H., & Bucher, B. Pupils' self-scored arithmetic performance: A practical procedure for maintaining accuracy. Journal of Applied Behavior Analysis, 1978, 11, 304.
- Johnson, V.S. Behavior modification in the correctional setting. Criminal Justice and Behavior, 1977, 4, 387-428.
- Kaufman, K.F., & O'Leary, K.D. Reward, cost and self-evaluation procedures for disruptive adolescents in a psychiatric hospital school. Journal of Applied Behavior Analysis, 1972, 5, 293-309.

- Kazdin, A.E., and Bootzin, R.R. The token economy: an evaluative review. Journal of Applied Behavior Analysis, 1972, 5, 343-372.
- Knapczyk, D.R., & Livingston, G. Self-recording and student teacher supervision: Variables within a token economy structure. Journal of Applied Behavior Analysis, 1973, 6, 481-486.
- Layne, C.C., Rickard, H.C., Jones, M.T., and Lyman, R.D. Accuracy of self-monitoring on a variable ratio schedule of observer verification. Behavior Therapy, 1976, 7, 481-488.
- Marholin, D., Siegel, L.J., & Phillips, D. Transfer and treatment: A search for empirical procedures. In M. Hersen, R.M. Eisler, & P.M. Miller (Eds.), Progress in behavior modification (Vol. 3). New York: Academic, 1976.
- Martin, G., & Pear, J. Behavior Modification: What it is and how to do it. Englewood Cliffs, N.J: Prentice-Hall, 1978.
- McCombs, D., Filipczak, J., Friedman, R.M., & Wodarski, J.S. Long-term follow-up of behavior modification with high risk adolescents. Criminal Justice and Behavior, 1978, 5(1), 21-34.
- McNees, M.P., Egli, D.S., Marshall, R.S., Schnelle, J.F., & Risley, T.R. Shoplifting prevention: Providing information through signs. Journal of Applied Behavior Analysis, 1976, 9, 399-405.
- Milan, M.A., & McKee, J.M. Behavior modification: Principles and applications in corrections. In D. Glaser (Ed.), Handbook of criminology. Chicago: Rand McNally, 1974.
- Millenson, J.R., and Leslie, J.C. Principles of behavior analysis. New York: Macmillan, 1979.
- Minkin, N., Braukmann, C.J., Minkin, B.L., Timbers, G.D., Timbers, B.I., Fixsen, D.L., Phillips, E.L., & Wolf, M.M. The social validation and training of conversation skills. Journal of Applied Behavior Analysis, 1976, 9, 127-140.
- Morris, E.K. A brief review of legal deviance: References in behavior analysis and delinquency. In D. Marholin (Ed.), Child behavior therapy. New York: Gardner, 1978.
- Nelson, R.O., Lipinski, D.P., and Boykin, R.A. The effects of self-recorders' training and the obtrusiveness of the self-recording device on the accuracy and reactivity of self-monitoring. Behavior Therapy, 1978, 9, 200-208.

- O'Leary, K.D., and Drabman, R. Token reinforcement programs in the classroom: A review. Psychological Bulletin, 1971, 75, 379-398.
- Parsonson, B.S. The modification of delinquent behaviors. Australian and NZ Journal of Criminology, 1972, 5(1), 49-56.
- Peacock, R., Lyman, R.D., and Rickard, H.C. Correspondence between self-report and observer-report as a function of task difficulty. Behavior Therapy, 1978, 9, 578-583.
- Pierce, C.H., & Risley, T.R. Recreation as a reinforcer: Increasing membership and decreasing disruptions in an urban recreation center. Journal of Applied Behavior Analysis, 1974, 7, 403-411.
- Sagotsky, G., Patterson, C.J., & Lepper, M.R. Training children's self-control: A field experiment in self-monitoring and goal-setting in the classroom. Journal of Experimental Child Psychology, 1978, 25, 242-253.
- Salzberg, C.L. Freedom and responsibility in an elementary school. In G. Semb (Ed.), Behavior analysis and education - 1972. Lawrence: University of Kansas, Support and Development Center for Follow Through, 1972.
- Santogrossi, D.A., O'Leary, K.D., Romanczyk, R.G., & Kaufman, K.F. Self-evaluation by adolescents in a psychiatric hospital school token program. Journal of Applied Behavior Analysis, 1973, 6, 277-287.
- Schnelle, J.F., Kirchner, R.E. McNees, M.P., & Lawler, J.M. Social evaluation research: The evaluation of two police patrolling strategies. Journal of Applied Behavior Analysis, 1975, 8, 353-365.
- Schwartz, A., & Goldiamond, I. Social casework: A behavioral approach. New York: Columbia University Press, 1975.
- Seymour, F.W., & Stokes, T.F. Self-recording in training girls to increase work and evoke staff praise in an institution for offenders. Journal of Applied Behavior Analysis, 1976, 9, 41-54.
- Stokes, T.F., & Baer, D.M. An implicit technology of generalization. Journal of Applied Behavior Analysis, 1977, 10, 349-367.
- Stumphauzer, J.S. (Ed.) Behavior therapy with delinquents. Springfield, IL: Charles C. Thomas, 1973.
- Stumphauzer, J.S. (Ed.) Progress in behavior therapy with delinquents. Springfield, IL: Charles C. Thomas, 1979.

- Tharp, R.G., & Wetzell, R.J. Behavior modification in the natural environment. New York: Academic, 1969.
- Turkewitz, H., O'Leary, K.D., & Ironsmith, M. Generalization and maintenance of appropriate behavior through self-control. Journal of Consulting and Clinical Psychology, 1975, 43, 577-583.
- Weir, K., & Ford, D. Modification of the behaviors of delinquent youths in a token economy: Lentora project. Aust. and NZ Journal of Criminology, 1977, 10, 153-164.
- Werner, J.S., Minkin, N., Minkin, B.L., Fixsen, D.L., Phillips, E.L., & Wolf, M.M. What should kids say to cops? An analysis of an "intervention program". Criminal Justice and Behavior, 1975, 2, 55-84.
- Willner, A.C., Braukmann, C.J., Kirigin, K.A., Fixsen, D.L., Phillips, E.L., & Wolf, M.M. The training and validation of youth-preferred social behaviors in child-care personnel. Journal of Applied Behavior Analysis, 1977, 10, 219-230.
- Willner, A.G., Braukman, C.J., Kirigin, K.A., & Wolf, M.M. Achievement Place: A Community treatment model for youths in trouble. In D. Marholin (Ed.), Child behavior therapy. New York: Gardner Press, 1978.
- Wolf, M.M. Social validity: The case for subjective measurement. Journal of Applied Behavior Analysis, 1978, 11, 203-214.
- Wood, R., and Flynn, J.M. A self-evaluation token system versus an external evaluation token system alone in a residential setting with predelinquent youths. Journal of Applied Behavior Analysis, 1978, 11, 503-512.